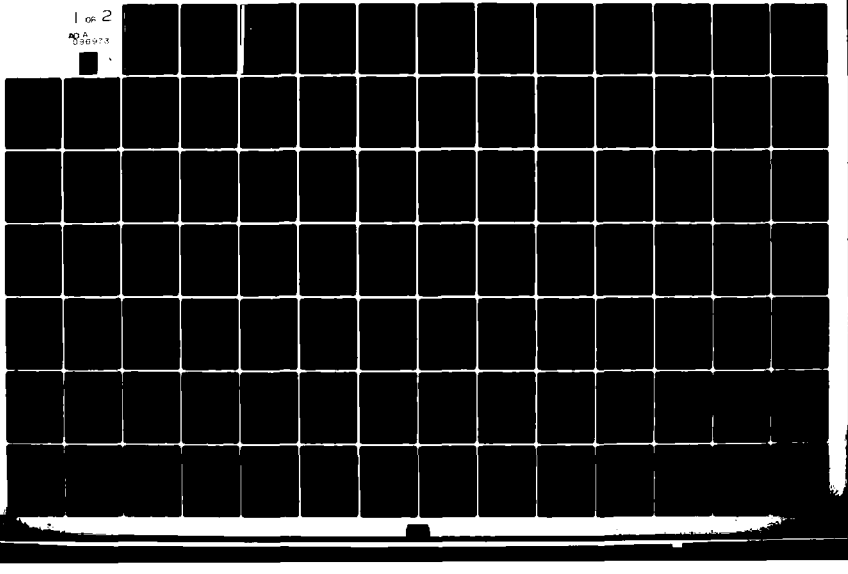
AD-A096 973 MASSACHUSETTS INST OF TECH CAMBRIDGE LAB FOR INFORMA--ETC F/G 12/1
TOPICS IN NONLINEAR FILTERING THEORY.(U)
DEC 80 D L OCONE AFOSR-77-3281
UNCLASSIFIED LIDS-TH-1058 AFOSR-TR-81-0232 NL

1 OF 2
AD A 096973



AD A096973

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER 19 AFOSR-TR-81-0232	2. GOVT ACCESSION NO. AD-A096973
4. TITLE (and Subtitle) 6 TOPICS IN NONLINEAR FILTERING THEORY	5. TYPE OF REPORT & PERIOD COVERED Interim technical
7. AUTHOR(s) 10 Daniel Ocone 15	8. CONTRACT OR GRANT NUMBER(s) AFOSR-77-3281
9. PERFORMING ORGANIZATION NAME AND ADDRESS Massachusetts Institute of Technology Lab for Information and Decision Systems Cambridge, MA 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 16 17 61102F 2304/A1
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332 11	12. REPORT DATE December 1980
	13. NUMBER OF PAGES 180 13) 12
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) 14) LIDS-TH-1057	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

1981

A

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This thesis studies two topics in the theory of nonlinear filtering, the use of multiple stochastic integrals to analyze filters, and the use of Lie algebraic and operator-theoretic techniqies to discover new, finite-dimensionally solvable filtering problems. The main results of the multiple integral techniqies are: (1.) A simplier and more insightful proof of a result of S. Marcus on filtering polynomials functions of a Gauss-Markov process. (2.) A formula for representing the product of two multiple integrals as a sum of

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DTIC FILE COPY

multiple integrals, thus providing a rudimentary calculus of multiple integral expansions.

3.) An expansion of the optimal means square filter as a ratio of two multiple integral expansions, a

4.) Integral equations for the kernels of the best mean square filter of the class of (finite) r order multiple integral expansions.

The problem of estimating a diffusion process observed in white noise is studied with Lie algebra techniques. Necessary conditions, and in the scalar case, necessary and sufficient conditions, are given for estimation algebra finite dimensionality. Examples of scalar problems with fin. dim. estimation algebras are discussed, and it is shown that, from among them, no new cases exist for which Zakai's equation can be solved by a Wei-Norman type method.

AFOSR-TR- 81-0232

December 1980

LIDS-TH-1058

Research Supported By:
Grant AFOSR-77-3281

TOPICS IN NONLINEAR FILTERING THEORY

Daniel Ocone

Laboratory for Information and Decision Systems
MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MASSACHUSETTS 02139

3 27

December 1980 LIDS-TH-1058

TOPICS IN NONLINEAR FILTERING THEORY

by

Daniel Ocone

This report is based on the unaltered thesis of Daniel Ocone submitted in partial fulfillment of the requirements for the degree of Doctor of Science at the Massachusetts Institute of Technology, June 1980. The research was conducted at the M.I.T. Laboratory for Information and Decision Systems, with support provided in part by grant AFOSR-77-3281[illegible].

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

TOPICS IN NONLINEAR FILTERING THEORY

by

DANIEL OCONE

Sc.B. Brown University
(1975)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF

DOCTOR OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1980

© Massachusetts Institute of Technology 1980

Signature of Author Daniel L. Ocone
Department of Mathematics
May 23, 1980

Certified by _____
Sanjoy K. Mitter
Thesis Advisor

Accepted by Michael Artin
Michael Artin
Chairman, Department Committee

2

TOPICS IN NONLINEAR FILTERING THEORY

by

DANIEL OCONE

Submitted to the Department of Mathematics
on May 23, 1980 in partial fulfillment of the
requirements for the Degree of Doctor of Science in
Mathematics

ABSTRACT

This thesis studies two topics in the theory of nonlinear filtering; the use of multiple stochastic integrals to analyze filters, and the use of Lie algebraic and operator-theoretic techniques to discover new, finite-dimensionally solvable filtering problems.

The main results of the multiple integral techniques are:

1. A simpler and more insightful proof of a result of S. Marcus on filtering polynomials functions of a Gauss-Markov process.

2. A formula for representing the product of two multiple integrals as a sum of multiple integrals, thus providing a rudimentary calculus of multiple integral expansions.

3. An expansion of the optimal mean square filter as a ratio of two multiple integral expansions.

4. Integral equations for the kernels of the best mean square filter of the class of (finite) r^{th} order multiple integral expansions.

The problem of estimating a diffusion process observed in white noise is studied with Lie algebra techniques. Necessary conditions, and in the scalar case, necessary and sufficient conditions, are given for estimation algebra finite dimensionality. Examples of scalar problems with fin. dim. estimation algebras are discussed, and it is shown that, from among them, no new cases exist for which Zakai's equation can be solved by a Wei-Norman type method.

Thesis Supervisor: Sanjoy K. Mitter
Title: Professor, Department of Electrical Engineering

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Sanjoy Mitter, for suggesting this research topic, for his continuous and inspiring involvement in its progress, and for the grant support of my last year (AFOSR-77-3281B).

I also want to thank the many fine researchers Prof. Mitter introdu ed me to--among others, Steve Marcus, V. Beneš, M. H. A. Davis, E. Wong, and John Baras. To them, through their helpful discussions with me, this thesis owes much.

I am further grateful to William Dunsmuir and Richard Dudley for being on my thesis committee and for reading and checking this technical work, to Nancy Toscano and company for their cheerful and professional typing, and to the NSF Graduate Fellowship Program for their financial support.

Finally, thanks are due to my family and friends for their support and encouragement over these long, graduate years.

TABLE OF CONTENTS

		Page
ABSTRACT		2
ACKNOWLEDGEMENTS		3
TABLE OF CONTENTS		4
CHAPTER 1	INTRODUCTION	6
	1.1 The Nonlinear Filtering Problem	6
	1.2 Summary of Thesis	9
	1.3 The Fundamental Problem; Mathematical Prerequisites	12
CHAPTER 2	MULTIPLE INTEGRALS	16
	2.1 Definition and Basic Properties of Multiple Integrals	16
	2.2 The Observation Semi-Martingale Case	24
	2.3 The Multiplication Formula	30
CHAPTER 3	ESTIMATION OF NONLINEAR FUNCTIONS OF GAUSSIAN PROCESSES	53
CHAPTER 4	MULTIPLE STOCHASTIC INTEGRALS AND NONLINEAR FILTERING	67
	4.1 Filter Expansions	69
	4.2 Best r^{th} Order Filters	80
	4.3 The Kalman Filter	90
	4.4 Quadratic Filters	95
CHAPTER 5	NONLINEAR FILTERING PROBLEMS WITH FINITE DIMENSIONAL ESTIMATION ALGEBRAS	109
	5.1 Estimation Algebras and Filtering	112

TABLE OF CONTENTS (CONTINUED)

	Page
5.2 Estimation of Absorbed Brownian Motion	128
5.3 Finite Dimensional Estimation Algebras.....	138
BIBLIOGRAPHY.....	170
APPENDIX 1 HOMOGENEOUS CHAOS THEORY.....	174
APPENDIX 2 SOME CONCEPTS FROM OPERATOR THEORY.....	177

CHAPTER 1: INTRODUCTION

1.1 The Nonlinear Filtering Problem

Nonlinear filtering theory is the study of a broad range of problems in the estimation of stochastic processes. A typical example concerns the estimation of a signal in additive noise. In this situation, one is interested in the properties of a stochastic process $\{x(t)|t\geq 0\}$ called the signal. One might want to know, for instance, the value of $f(x(t))$, for a function f, or even the value $g(x(s);0\leq s\leq t)$ for a functional of the past of $x(\cdot)$ up to time t. However, information about $x(\cdot)$ is available only through observation of the process

$$h(t) = \int_0^t h(s,x(s))ds + w(t) \qquad (1.1)$$

in which h is a given function and $w(t)$ is a "noise", usually an independent increments process. Thus estimates, or rather, filters, for $f(x(t))$ and $g(x(s);0\leq s\leq t)$ must be constructed from $\{y(s)|0\leq s\leq t\}$.

Minimization of the mean square filtering error is the criterion generally chosen to guide filter design. Thus, in linear filtering theory the goal is to produce the best (mean square) estimate that is a linear functional of $\{y(s)|0\leq s\leq t\}$. Nonlinear filtering theory goes further; it asks for the best mean square estimate given the past of $y(\cdot)$. If $Ef^2(x(t))<\infty$ and if $\mathcal{F}_t^y$ denotes the

σ -algebra $\sigma\{y(s)|0\leq s\leq t\}$, it is well known that this estimate is $\pi_t(f) = E\{f(x(t))|F_t^y\}$. The goal of nonlinear filtering is to compute or to approximate this conditional expectation.

Interest in filtering problems stems from their central role in several applied subjects. For example, in the theory of communication (Van Trees [48]), (1.1) is a common model for a signal sent in a noisy channel; successful transmission of information requires extracting the signal from the noise. It may also be necessary to decide on the basis of $\{y(s)|s\leq t\}$ between two possibilities, $h(s,x(s))\equiv 0$, $s\leq t$, or $h(s,x(s))$ = a given signal. This "signal detection" problem is closely related to optimal filtering (E. Wong [46]). Stochastic control problems, in which a control is to be chosen so as to influence signal process behavior, can also involve filtering if the control is allowed to depend on noisy or partial observations of the signal (see Fleming and Rishel [14] and references cited therein.).

The modern literature of nonlinear filtering begins with the contribution of Kalman and Bucy [24], who formulated and solved the model (1.1) for the case in which $x(t)$ is a Gaussian diffusion, $h(s,x)$ is a linear in x, $w(t)$ is Brownian, and $f(x(t))=x(t)$. Their main result, to be stated in lemma 3.1, proves that the conditional density of $x(t)$ given F_t^y is Gaussian and provides a method to compute the conditional mean and covariance recursively.

for few other cases is such a complete and easily constructed solution available. However, two very powerful characterizations of optimal filters are known to hold in quite general situations. The first is a Bayes-type formula for $\pi_t(f)$, which is due to Kallianpur and Striebel [22] and which, in essence, represents $\pi_t(f)$ by a functional integration in process path space (see Section 1.3). It is valid for Brownian noise $w(t)$ with minimal restrictions on $x(\cdot)$, h, and f. When the signal is Markovian, $\pi_t(f)$ can be further characterized as the solution of a stochastic differential equation (Fujisaki, Kallianpur and Kunita [15]). In general, $\pi_t(f)$ cannot be found from this result because the coefficients of the filter equation involve optimal estimates of $\phi(x(t))$ for functions ϕ different from f. Thus additional equations are required to compute $\pi_t(\phi)$, which in turn require estimates of yet other functions of $x(t)$. The resulting system of equations is in general infinite-dimensional. The cleanest formulation of this infinite dimensionality is Zakai's [47] stochastic partial differential equation for an unnormalized version of the conditional density, assuming this density exists (see Chapter 5). Finally, several very recent developments promise new insights. V. Benes [1] has derived new examples of explicitly solvable filtering problems, and Brockett and Clark [17], Brockett [5], and Mitter [35] have begun applying Lie algebraic and operator techniques to the study of conditional

density equations. These developments will be discussed in Chapter 5.

The above,briefly outlined results constitute the principal highlights of nonlinear filtering theory, but, despite their mathematical depth, they remain incompletely developed. For many common filtering problems little is actually known about the filter structure and one must resort to reasonable, but ad hoc techniques. A powerful and general theory for building, analyzing and comparing suboptimal designs does not exist.

1.2 Summary of Thesis

This thesis studies two different ideas for analyzing nonlinear filtering problems. The first is that of evaluating or approximating filters by expansions in series of multiple stochastic integrals. Such an approach is motivated by the fact that the optimal estimate $\pi_t(f)$ may be thought of as a functional $F_t(y(\cdot))$ of the observation process. It is then possible to explore $\pi_t(f)$ within the framework of a representation theory for F, for instance, one that expands F in a series of simpler and more easily manipulated basis functionals. Multiple integrals are ideally suited for this, because they are easy to handle and because they can represent a large class of functionals F

(see theorem 2.2).

The second idea differs from the first in method and style. Rather than expansion or approximation, it studies the question of when a filtering problem can be solved in an exact, finite dimensional manner. And rather than being probabilistic, the techniques are algebraic and operator-theoretic. Brockett and Clark [7], Brockett [5,6] and Mitter [35] have shown recently that certain Lie algebras of operators, called estimation algebras, can be associated to the problem of filtering a Markov process observed in white noise. In examples with known, finite dimensionally computable conditional densities, that is, the examples of Kalman and Bucy [24] and of Benes [1], the estimation algebra is also finite dimensional. Conversely, it is widely conjectured that given appropriate hypotheses, Lie algebra finite dimensionality will imply the existence of a finite dimensionally computable expression for the conditional density. This suggests the strategy taken up in the second part of the thesis research: seek all problems with finite dimensional estimation algebras and try to solve them.

The main results of our investigation are presented in the following chapter by chapter summary of the thesis. Chapter 2 defines the multiple stochastic integral and develops some of its fundamental properties. The main result here is the multipli-

cation formula, (theorem 2.4, Section 2.3), which describes how to re-expand the product of two multiple integrals as a sum of multiple integrals and which is an important calculational and conceptual tool in the theory of Chapter 4. Technical lemmas needed in Chapters 3 and 4 are also stated and proved.

Chapter 3 contains a proof of a result originally due to S. Marcus [29] on the finite dimensional solvability of filters for estimating polynomial functions of a Gauss-Markov signal process given linear, but noisy observations. The proof here sets the problem in the context of Gaussian process theory by using multiple integrals and homogeneous chaos theory. It is simpler than Marcus' original proof and explains more clearly how and why a finite number of statistics characterize the optimal estimate. This work was done jointly with S. Marcus and S. K. Mitter.

In Chapter 4 we present expansion theories for the general filtering model of estimating a signal in white Gaussian noise. First, we derive a representation of the full optimal filter as a ratio of multiple integral expansions. In effect, this representation evaluates the functional integrals of the Kallianpur-Striebel formula with multiple integrals. Secondly, we pose a basic problem, suggested by the multiple integral idea, for the design of suboptimal filters: For any r, what is the best (mean square) estimator having the form of an r^{th} order multiple

integral expansion? Using the expansion representation in conjunction with the multiplication formula, we derive integral equations for the kernels of a best r^{th} order estimate. We then rederive the Kalman-Bucy filter and discuss the case r=2 as examples of the technique.

Chapter 5 discusses the Lie algebra approach to finite dimensional filter computation. The main results are presented in Section 5.3. For vector diffusion signals with non-singular, constant local covariance, a fairly restrictive necessary condition is given for estimation algebra finite dimensionality. In the scalar case, this allows all possible problems with finite dimensional estimation algebras to be listed. A solution of some of these filtering problems is then attempted using a method developed and discussed in Sections 5.1 and 5.2. The result is that only those previously known examples of Benes can be solved finite-dimensionally by this method.

It is worth remarking that the last chapter is discursive in style and does not present a complete theory. This chapter is a preliminary report and discussion on calculations of interest to a new, developing theory with important implications. To shorten the exposition and concentrate on the main idea, we have omitted certain cases from the analysis, but, as shall be mentioned, the

results extend formally to them. In this regard, work remains to be done in building more complete results. However, we feel the ideas are sufficiently well-developed and interesting to merit inclusion.

1.3 The Fundamental Problem: Mathematical Prerequisites

The work of this thesis employs techniques from a variety of fields. Chapters 2-4 assume familiarity with certain elements of stochastic process theory, in particular, stochastic integration with respect to Brownian motion, stochastic differential equations, and Ito's rule. Chapter 3 uses some homogeneous chaos theory, which is summarized briefly in Appendix 1. Finally, Chapter 5 requires familiarity with the use of Lie algebra/Lie group methods in systems theory and with the theory of self-adjoint operators. Appendix 2 states the basic definitions and results that are needed from operator theory.

We will adopt the following conventions throughout the thesis: all Brownian motions are assumed to have mean zero and unity scale; if $\{z(t)|t\geq 0\}$ is a stochastic process, $F_t^z = \sigma\{z(s)|s\leq t\}$ denotes the σ-algebra generated by $z(s)$ for $s\leq t$.

We now state the precise filtering problem to be considered in the thesis. Let $\{x(t) | t \in [0,T]\}$ be a measurable real-valued process on a probability space (Ω,F,P). Let $h(s,x)$ be a Borel measurable function. Set

$$y(t) = \int_0^t h(s,x(s))ds + w(t) \qquad t \in [0,T] \qquad (1.2)$$

and assume

(i) w is a Brownian motion independent of the signal process $x(\ )$

(ii) $E \int_0^T h^2(s,x(s))ds < \infty$

Definition 1.1. A process $\{y(t) | t \in [0,T]\}$ defined by 1.2 satisfying the stated assumptions is called an observation semimargingale. Given a functional $f(t;x(s),s\leq t)$ of the past of $x(\cdot)$, we want to compute the optimal mean square estimate

$$\pi_t(f) = E\{f(t;x(s),s\leq) | F_t^y\} .$$

The following theorem of Kallianpur and Striebel [22] will be a principle theoretical tool of this thesis. For a good exposition, see Wong [46].

Theorem 1.1 (Kallianpur, Striebel). Let

$$\frac{dP_0}{dP} = \exp\left[-\int_0^T h(x(s))dw(s) - \frac{1}{2}\int_0^T h^2(x(s))ds\right].$$

Then (i) P_0 is a probability measure, P and P_0 are mutually absolutely continuous, and

$$\frac{dP}{dP_0} = \exp\left[\int_0^T h(x(s))dy(s) - \frac{1}{2}\int_0^T h^2(x(s))ds\right]$$

(ii) $E_0\{\frac{dP}{dP_0}|F_t^{x,y}\} = \exp\left[\int_0^t h(x(s))dy(s) - \frac{1}{2}\int_0^t h^2(x(s))ds\right]$.

(iii) W.r.t. P_0, $y(\cdot)$ is a Brownian motion independent of $x(\cdot)$.

(iv) $x(\cdot)$ has the same law w.r.t. P_0 as w.r.t. P.

(v) $E\{f(t;x(s),s\leq t)|F_t^y\}$

$$= \frac{E_0\{f(t;x(s),s\leq t)\frac{dP}{dP_0}|F_t^y\}}{E_0\{\frac{dP}{dP_0}|F_t^y\}} \quad .$$

Finally, the concept of innovations will occasionally be needed.

Definition 1.2. The innovations process associated to the filtering problem of (1.2) is

$$\nu(t) = y(t) - \int_0^t \pi_s\,(h(s,x(s))\,ds$$

Interestingly, given mild restrictions on the nature of $x(\cdot)$ and $h(s,x)$, $\nu(t)$ is a Brownian motion (Lipster and Shiryayev [28]).

CHAPTER 2 MULTIPLE INTEGRALS

This chapter will define multiple stochastic integrals with deterministic kernels, discuss their basic properties, and establish both theoretical and technical results that are important in filtering applications. The main result of this chapter is the multiplication formula of theorem 2.4 in section 3.

2.1 Definition and Basic Properties of Multiple Integrals

This section is devoted to a brief exposition of the multiple Wiener integral and its elementary properties. Most of the material is well known and is due to Ito (20), who developed the definition in its present form and demonstrated its connection to homogeneous chaos theory. In addition, we prove some technical results, including a construction to produce multiple integrals recursively from stochastic differential equations, important in subsequent work.

Let $\{b(t)\}_{t\geq 0}$ be a Brownian motion, and let $F_t \equiv \sigma\{b(s) | s \leq t\}$ denote its associated family of sub-σ-algebras. If $\phi(s,\omega)$ is a measurable random process adapted to F_t, (i.e., $\phi(t,\cdot)$ is F_t-measurable for every t), and if $E\int_0^T \phi^2(s)\,ds < \infty$, then for $t \leq T$ we can define the measurable, adapted process

$$\int_0^t \phi(s)db(s);$$

see, for instance, Liptser and Shiryayev [28]. Recall the properties of this integral

$$E\left[\int_0^t \phi(s)db(s)\right] = 0$$

$$E\int_0^t \phi(s)db(s)\int_0^t \psi(s)db(s) = \int_0^t E\phi(s)\psi(s)ds \quad . \tag{2.1}$$

We will use this single integral to define multiple integrals by iteration, a technique different than Ito's, but equivalent in result up to a multiplicative constant.

Definition 2.1:

(i) $f \in L^2([0,T]^r)$ is *separable on* $s_1 \geq s_2 \geq \cdots \geq s_r$ if

$f(s_1,\cdots,s_r) = \sum_{i=1}^{N} \gamma_1^{(i)}(s_1)\cdots\gamma_r^{(i)}(s_r)$ for

$T \geq s_1 \geq s_2 \geq \cdots \geq s_r$.

(ii) $\hat{L}_s^2([0,T]^r) = \{f \in L^2([0,T]^r) \mid f$ is

(a) separable on $s_1 \geq \cdots \geq s_r$, and

(b) symmetric$\}$

(iii) $\hat{L}^2([0,T]^r) = \{f \in L^2[0,T]^r \mid f$ is symmetric$\}$.

Definition 2.2 Let $f \in \hat{L}^2_s([0,T]^r)$, $f = \gamma_1(s_1)\cdots\gamma_r(s_r)$ on $s_1 \geq \cdots \geq s_r$. The *r^{th} multiple stochastic integral* of f up to time t is defined inductively by

$$I^r_t(f) = \int_0^t \gamma_1(s)\, I^{r-1}_s(\gamma_2\cdots\gamma_r)db(s) \tag{2.2}$$

and $I^r_t(f)$ is defined on all of $\hat{L}^2_s([0,T]^r)$ by linear extension. Also, we adopt the convention, $I^0_T(f) = f$.

Note that (2.2) is meaningful, because at each step, using separability and induction, $\gamma_1(s)\, I^{r-1}_s(\gamma_2\cdots\gamma_r)$ is a measurable, F_t-adapted process and hence may be integrated.

Theorem 2.1 For $f, g \in \hat{L}^2_s([0,T]^r)$

$$E\, I^r_t(f) = 0 \tag{2.3}$$

$$E\, I^r_t(f)\, I^r_t(g) = \frac{1}{r!}\,(f,g)_{L^2} \tag{2.4}$$
$$= \int_0^t\int_0^{s_1}\cdots\int_0^{s_r} f(s_1,\cdots,s_r)g(s_1,\cdots,s_r)ds_r\cdots ds_1$$

Therefore, $\sqrt{r!}\, I^r_t$ is an isometry between $\hat{L}^2_s([0,T]^r)$ and $I^r_t[\hat{L}^2_s([0,T]^r)]$. Since $\hat{L}^2_s([0,T]^r)$ is dense in $\hat{L}^2([0,T]^r)$ we can extend the definition of I^r_t to $\hat{L}^2([0,T]^r)$ by continuity.

Proof By (2.1), (2.3) and (2.4) certainly hold for $r = 1$. Let $f = \gamma_1(s_1)\cdots\gamma_r(s_r)$, $g = \alpha_1(s_1)\cdots\alpha_r(s_r)$, and suppose (2.3) and (2.4) hold for $r - 1$. Then

$$E\, I_t^r(f) = E \int_0^t \gamma_1(s)\, I_s^{r-1}(\gamma_2\cdots\gamma_r)ds =$$

$$\int_0^t \gamma_1(s)\, E\, I_s^{r-1}(\gamma_2\cdots\gamma_r)ds = 0$$

$$E\, I_t^r(f)\, I_t^r(g) = \int_0^t \gamma_1(s)\alpha_1(s)\, E\, I_s^{r-1}(\gamma_2\cdots\gamma_r)\, I_s^{r-1}(\alpha_2\cdots\alpha_r)ds$$

$$= \int_0^t \cdots \int_0^{s_{r-1}} \gamma_1(s_1)\cdots\gamma_1(s_r)\alpha_1(s_1)\cdots$$

$$\alpha_r(s_r)ds_r\cdots ds_1.$$

The theorem follows by induction on r.

Remarks 1 By continuity, (2.3) and (2.4) hold for all f in $\hat{L}^2([0,T]^r)$.

2. It is not necessary to require that f be symmetric since integration is carried out only over the set $s_1 \geq s_2 \geq \cdots \geq s_r$. However, the convention of symmetry is useful later on.

3. For $f \in \hat{L}^2[[0,T]^r]$, let $f(s,\ldots)$ denote the section of f at s. We want to write

$$I_t^r(f) = \int_0^t I_s^r(f(s,\ldots))ds, \tag{2.5}$$

but, to do this, we need a measurable version of $I_s^r(f(s,\ldots))$. If f is separable on $s_1 \geq s_2 \geq \cdots \geq s_r$ this measurable version is immediately guaranteed; indeed, this is how we defined I_t^r. If f is not separable, let $\{f_n\}$ be a sequence of separable functions such that $||f_n - f||_{L^2} \to 0$. We then see that $\lim_{n\to\infty}$ m.s $I_s^r(f_n(s,\cdots))$ is a measurable version of $I_s^r(f(s,\cdots))$ and hence (2.5) is valid.

Let $\bigoplus_{r=0}^{\infty} \overline{G}_r = L^2(\Omega,F_T,P)$ be the homogeneous chaos decomposition of $\{b(s)|s\leq T\}$, (see Appendix 1 for the definition of this decomposition).

Theorem 2.2 (Ito)

For every r, $\overline{G}_r = \{I_T^r(f)|f \in \hat{L}^2([0,T]^r)\}$. Thus, if $\phi \in L^2(\Omega,F_T,P)$, there exist kernels $k_r \in \hat{L}^2([0,T]^r)$ such that

$$\phi = k_0 + \sum_{r=1}^{\infty} I_T^r(k_r).$$

Proof The full proof may be found in Ito (20), but let us sketch briefly why it is true. First, note that the spaces $V_r = I_T^r(\hat{L}^2([0,T]^r))$ are closed. Moreover, by using step functions we can see that $V_r \subset P_r$ and $P_r \subset V_0 + \cdots + V_r$, $(V_0 = \mathbb{R})$, (see appendix 1 for the definitions of P_r and $\bar{P}_r$). Since $\bar{G}_r = P_r \ominus P_{r-1}$, it suffices to show $V_r \perp V_q$ for $r \neq q$, that is, that integrals of different order are orthogonal. This fact, one of the salient features of multiple Wiener integrals, is easily verified. If $f \in \hat{L}^2([0,T]^r)$, $g \in \hat{L}^2([0,T]^q)$, $r > q$, then, from (2.5) and (2.1),

$$E\, I_T^r(f)\, I_T^q(g) = \int_0^T E\, I_s^{r-1}(f(s,\ldots)) I_s^{q-1}(g(s,\ldots))ds$$

$$= \cdots$$

$$= \int_0^T \cdots \int_0^{s_{q-1}} E\, I_{s_q}^{r-q}[f(s_1,\cdots,s_q,\cdots)]g(s_1,\cdots,s_q)$$

$$ds_q \cdots ds_1$$

$$= 0 .$$

Multiple stochastic integrals generalize easily to the vector case. If $b^T(t) = (b_1(t),\cdots,b_\nu(t))$ is a ν-dimensional Brownian motion and if $f \in \hat{L}^2([0,T]^r)$, we can define

$$I_T^{r;\ell_1\cdots\ell_r}(f) = \int_0^t \cdots \int_0^{s_{r-1}} f(s_1,\cdots,s_r)db_{\ell_r}(s_r)\cdots db_{\ell_1}(s_1) .$$

If $(r,\ell_1\cdots\ell_r) \neq (q,k_1\cdots k_q)$, then the integrals $I_t^{r;\ell_1,\cdots,\ell_r}(f)$ and $I_t^{q;\ell_1,\cdots,\ell_q}(g)$ are orthogonal.

In the case of separable kernels, a construction observed by Brockett (3) for realizing deterministic Volterra series can be adapted to produce multiple stochastic integrals from stochastic differential equations. This result motivates the use of multiple integrals because it says we can calculate, or at least approximate them recursively. Moreover, the criterion of kernel separability is used in Chapter 3 to prove finite dimensional computability of certain optimal filters.

Theorem 2.3. Let $f \in \hat{L}_s^2([0,T]^r)$. Then, for some n, there exists an R^n-valued process $z(t)$ that satisfies

$$dz(t) = \sum_{\ell=1}^{\nu} A_\ell(t)z(t)db_\ell(t) \quad z(0) = z_0$$

for some $n \times n$ matrix functions $A_\ell(t)$, $\ell = 1,\cdots, \nu$, and for some n-vector function $c(t)$, such that $I_t^{r;\ell_1,\cdots,\ell_r}(f) = c^T(t)z(t), t\leq T$.

Proof It suffices to consider $f = \gamma_1(s_1)\cdots\gamma_r(s_r)$. Suppose $\ell_{i_1} = \cdots \ell_{i_j} = \ell$, $i_1 < i_2 < \cdots < i_j$, and define the $(r+1)\times(r+1)$ matrix $A_\ell(t)$

$$A_\ell^{ij}(t) = \begin{cases} \gamma_{i_k}(t), (i,j) = (i_k, j_{k+1}) \; 1 \le k \le j \\ 0 \;, \text{ otherwise} \end{cases}$$

That is:

$$A_\ell(t) = \begin{matrix} \\ i_1\text{-row} \\ \\ i_j\text{-row} \\ \\ \end{matrix} \begin{bmatrix} 0 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \gamma_{i_1}(t) & & & \\ & & & \ddots & \ddots & & \\ & & & & \ddots & \ddots & \\ & & & & & 0 & \gamma_{i_j}(t) \\ & & & & & & \ddots \\ & & & & & & 0 \end{bmatrix}$$

Otherwise, define $A_\ell(t) \equiv 0$. Consider the system $z(t) \in \mathbb{R}^{r+1}$

$$dz(t) = \sum_{\ell=1}^{\nu} A_\ell(t) z(t) d\nu_\ell(t), \; z^T(0) = (0, \cdots, 0, 1) \; .$$

We have

$$z_{r+1}(t) \equiv 1$$

$$z_r(t) = \int_0^t \gamma_r(s) db_{\ell_r}(s)$$

$$\vdots$$

$$z_1(t) = \int_0^t \cdots \int_0^{s_{r-1}} \gamma_1(s_1)\cdots\gamma_r(s_r)db_{\ell_r}(s_r)\cdots db_{\ell_1}(s_1).$$

Thus $I^{r;\ell_1\cdots\ell_r}(f) = (1,0,\cdots,0)\, z(t)$.

Finally, we will need a Fubini-type lemma on the interchange between ds and db integrations

Lemma 2.1 Let $f \in \hat{L}^2([0,T]^r)$. For $t \leq T$

$$\int_0^t I_s^{r-1}(f(s,\cdots)ds = \int_0^t \cdots \int_0^{s_{r-2}} \int_{s_1}^t f(u,s_1,\cdots,s_{r-1})$$

$$du\, db(s_{r-1})\cdots db(s_1) \quad . \tag{2.6}$$

Proof Define $g_t(s_1,\cdots,s_{r-1}) = \int_{s_1}^t f(u,s_1,\cdots,s_{r-1})du$. The r.h.s. of (2.6) is $I_t^{r-1}(g_t)$. To prove the lemma, simply verify that

$$E[\int_0^t I_s^{r-1}(f(s,\cdots))ds - I_t^{r-1}(g_t)]^2 = 0$$

by using the basic properties of the multiple stochastic integral.

2.2 The Observation Semi-Martingale Case

For purposes of filtering we must define multiple stochastic integrals

$$\int_0^t \cdots \int_0^{s_{r-1}} f(s_1,\cdots,s_r)dy(s_r)\cdots dy(s_1) \qquad (2.7)$$

with respect to observation semi-martingales

$$y(t) = \int_0^t x(s)ds + w(t) \quad .^{*} \qquad (2.8)$$

(Recall, from definition 1.1 of observation semi-martingales, that $x(\cdot)$ and $w(\cdot)$ are assumed independent, w is Brownian, and $E\int_0^T x^2(s)ds < \infty$ for some T, $0 < T < \infty$.) Such integrals are well known and are developed extensively in martingale theory; Meyer [33] is the best reference. However, the structure of the observation semi-martingale case allows a simple construction, which we develop here. Begin by noticing that, as stated in Theorem 1.1, $y(\cdot)$ is mutually absolutely continuous w.r.t. Brownian motion; if P is the original measure under which the processes of (2.8) are defined, there exists a P_0 mutually absolutely continuous w.r.t. P, such that $y(\ )$ is Brownian on (Ω,F,P_0) for $t \leq T$. Therefore, for $f \in L^2([0,T]^r)$ we define (2.7) as the multiple Wiener integral of the previous section by working on the measure space (Ω,F,P_0) and we call this integral $I_t^r(f)$ without reference to measure.

Remark The process with respect to which multiple integrals are taken will always be clear from context and so will not be indicated in the notation $I_t^r(f)$.

---

*For simplicity of notation, we have set $h(s,x)= \bar{x}$ in (2.8) (see section 1.3). The results to follow are valid for general h satisfying the conditions specified in section 1.3.

For F_T^y -- adapted processes $\phi(s,w)$ such that $P\{\int_0^T \phi^2(s)ds < \infty\} = 1$, we can define the integral

$$\int_0^t \phi(s)dy(s) \equiv \int_0^t \phi(s)x_s ds + \int_0^t \phi(s)dw(s) \qquad (2.9)$$

(see Liptser and Shiryayev [28]). As with the Brownian case, (2.7) may be interpreted as an iteration of (2.9)

Lemma 2.2 Let $f \in \hat{L}^2([0,T]^r)$.

$$I_t^r(f) = \int_0^t I_s^{r-1}(f(s,\cdots))dy(s) \qquad t \leq T.$$

Proof: This result is an easy consequence of the more general fact: the process $\int_0^t \phi(s)dy(s)$ defined in (2.9) is stochastically equivalent to the process $(\int_0^t \phi(s)dy(s))_{P_0}$ formed by working on (Ω,F,P_0) where y is Brownian. The equivalence of these integrals is obvious for stochastic step functions

$$\phi(s,\omega) = \sum_{i=1}^n \phi(t_i,\omega)1_{(t_i,t_{i+1}]}(s) \qquad *$$

and it follows for the general case by taking limits of such step functions.

---

* $1_{(t_i,t_{i+1}]}(s)$ = indicator function of $(t_i,t_{i+1}]$.

The expressions (2.3) and (2.4) for the mean and covariance of multiple integrals no longer holds in the semi-martingale case. However, it is important in later calculations to evaluate these moments, and, for this, the next lemma is useful.

Lemma 2.3 Suppose $E[\int_0^T x^2(s)ds]^r < \infty$. Then for $k \leq r$ and $f \in \hat{L}^2([0,T]^k)$

(i) $E[I_t^k(f)]^2 \leq M_k ||f||^2_{L^2}$; $M_k < \infty$ is independent of f

(ii) $E\, I_t^k(f) = \int_0^t \cdots \int_0^{s_{k-1}} f(s_1,\cdots,s_k)$

$$E\, x(s_1)\cdots x(s_k)ds_k\cdots ds_1 .$$

Proof. We will actually prove by induction the more general result: for $r \geq \ell \geq k$ $\sigma_k,\cdots,\sigma_\ell \in [0,T]$

$$E[x(\sigma_\ell)\cdots x(\sigma_{k+1})I^k_{\sigma_k}(f)]^2 \leq h_{\ell,k}(\sigma_{k+1},\cdots,\sigma_\ell)||f||^2 \quad (2.10)$$

where $h_{\ell,k} \in L^1([0,T]^{\ell-k})$, and

$$E[x(\sigma_\ell)\cdots x(\sigma_{k+1})I^k_{\sigma_k}(f)] = \int_0^{\sigma_k}\int_0^{s_1}\cdots\int_0^{s_{k-1}} f(s_1,\cdots,s_k) \quad (2.11)$$

$$E[x(s_1)\cdots x(s_k)\,(\sigma_{k+1})\cdots x(\sigma_\ell)]ds_k\cdots ds_1$$

Lemma 2.3 is the case $\ell = k$ for every $k \le r$. First we demonstrate (2.10) and (2.11) for $r \ge \ell \ge k = 1$, using the iterative formula of lemma 2.2 and the independence of $x(\cdot)$ and $w(\cdot)$. Thus

$$E[x(\sigma_\ell)\cdots x(\sigma_2)\int_0^{\sigma_1} f(s)dy(s)]^2 = E[x(\sigma_\ell)\cdots x(\sigma_2)\{\int_0^{\sigma_1} f(s)x(s)ds + \int_0^{\sigma_1} f(s)dw(s)\}]^2 \le [2\varepsilon \int_0^T [x(\sigma_\ell)\cdots x(s)]^2 ds + 2\, E[x(\sigma_\ell)^2\cdots x(\sigma_2)^2]]\|f\|^2 = h_{\ell,1}(\sigma_2,\cdots,\sigma_\ell)\|f\|^2 . \tag{2.12}$$

To derive the inequality in (2.12), the Cauchy-Schwarz inequality is used several times. $h_{\ell,1} \in L^1([0,T]^{\ell-1})$ for $\ell \le r$ because $E[\int_0^T x^2(s)ds]^r < \infty$. Likewise

$$E[x(\sigma_\ell)\cdots x(\sigma_2)\int_0^{\sigma_1} f(s)dy_s] = E[x(\sigma_\ell)\cdots x(\sigma_2)[\int_0^{\sigma_1} f(s)x(s)ds + \int_0^{\sigma_1} f(s)dw(s)]] = \int_0^{\sigma_1} f(s)E[x(s)x(\sigma_2)\cdots x(\sigma_\ell)]ds. \tag{2.13}$$

Now suppose (2.10) and (2.11) are true for a fixed k and all ℓ, $r \ge \ell \ge k$. Again, using $I_t^{k+1}(f) = \int_0^t I_s^k(f(s,\cdot))dy(s)$, Cauchy-Schwarz, and induction,

$$E[x(\sigma_\ell)\cdots x(\sigma_{k+2})I_{\sigma_{k+1}}^{k+1}(f)]^2 \le$$

$$2\int_0^{\sigma_{k+1}}\int_0^{s_1} E[x(\sigma_\ell)\cdots x(\sigma_{k+2})x(s_1)I_{s_2}^{k}(f(s_2,\cdots))]^2 ds_2 ds_1$$

$$+ 2\int_0^{\sigma_k} E[x(\sigma_\ell)\cdots x(\sigma_{k+2})I_s^k(f(s,\cdots))]^2 ds$$

$$\le [2\int_0^T h_{\ell,k}(s,\sigma_{k+2},\cdots,\sigma_\ell)ds + 2h_{\ell-1,k}(\sigma_{k+2},\cdots,\sigma_\ell)]\|f\|^2$$

$$= h_{\ell,k+1}(\sigma_{k+2},\cdots,\sigma_\ell)\|f\|^2.$$

By induction, $h_{\ell,k+1} \in L^1([0,T]^{\ell-k-1})$. Thus (2.10) is true for $k+1$. That (2.10) holds for k also implies

$$E\int_0^T I_s^k(f(s,\cdots))ds < \infty \quad .$$

Thus, because of (2.3),

$$E\int_0^t I_s^k(f(s,\cdots))dw(s) = 0, \quad \text{for } t \le T.$$

With the aid of this equality we can prove that (2.11) also is true for $k+1$.

This completes the induction step. Induction stops at $k = r$ since we have required $r \geq \ell \geq k$ in order to apply $E(\int_0^T x_s^2 ds)^r < \infty$.

2.3 The Multiplication Formula

To any given functional $\phi \in L^2(\Omega, F_t^b, P)$ of a Brownian motion $b(t)$, $t \leq T$, one can associate a sequence of kernels $\{k_r\}_{r=0}^{\infty}$, $k_r \in \hat{L}^2([0,T]^r)$, such that

$$\phi = k_0 + \sum_{n=1}^{\infty} I_t^r(k_n) \quad .$$

For applications, it is necessary to have rules for manipulating this representation, rules that describe how the kernels k_n behave under various transformations of the ϕ's. This section states, proves and discusses such a rule for the simplest case; if $f \in \hat{L}^2([0,T]^r)$, $g \in \hat{L}^2([0,T]^q)$, what are the kernels $\{\ell_i\}_{i=0}^{\infty}$ such that

$$I_t^r(f) I_t^q(g) = \ell_0 + \sum_{r=0}^{\infty} I_t^r(\ell_r) \quad .$$

The answer will require some new definitions.

Definition 2.3

(i) $P_r \equiv$ projection of $L^2([0,T]^r)$ onto $\hat{L}^2([0,T]^r)$:

$$(P_r h)(\sigma_1,\cdots,\sigma_r) = \frac{1}{r!}\sum_{\tau\in S_r} h(\sigma_{\tau(1)},\cdots,\sigma_{\tau(r)})$$

where S_r = permutation group on r letters.

(ii) For integers r,q,k, $0 \le k \le \min(r,q)$, and $f \in \hat{L}^2([0,T]^r)$, $g \in \hat{L}^2([0,T]^q)$

$$(f_k \otimes (t) g_k)(\sigma_1,\cdots,\sigma_{r+q-2k})$$

$$\equiv \frac{1}{k!}\int_0^t \cdots \int_0^t f(s_1,\cdots,s_k,\sigma_1,\cdots,\sigma_{r-k})$$

$$g(s_1,\cdots,s_k,\sigma_{r-k+1},\cdots,\sigma_{r+q-2k})\, ds_k\cdots ds_1$$

(iii) $f_k \odot (t)\, g_k(\sigma_1,\cdots,\sigma_{r+q-2k}) \equiv$

$$P_{r+q-2k}[f_k \otimes (t) g_k(\sigma_1,\cdots,\sigma_{r+q-2k})]$$

(iv) $f \odot g = f_0 \odot (t) g_0 =$

$$P_{r+q}[f(\sigma_1,\cdots,\sigma_r) g(\sigma_{r+1},\cdots,\sigma_{r+q})]$$

$\odot(t)$ is the basic operation by which new kernels are created from old, and, indeed, we will show in lemma 2.4 that $f_k \odot (t) g_k : \hat{L}^2([0,T]^r) \times \hat{L}^2([0,T]^q) \to \hat{L}^2([0,T]^{r+q-2k})$. To better under-

stand $\odot(t)$, it is useful to think of the functions f and g as tensors, which they in fact are under the isomorphism $L^2([0,T]^r) = L^2([0,T]) \otimes \cdots \otimes L^2([0,T])$ (r-fold). Then $f_k \otimes(t) g_k$ may be viewed as a tensor contraction, and $f_k \odot(t) g_k$ as a symmetrized tensor contraction. The notation $f_k \odot (t) g_k$ is meant to recall the summation notation, i.e., a sum (integral) is taken along the first k indices of f and g. It is in this definition that we make use of the convention that the integrands f and g are symmetric; otherwise $\otimes(t)$ would have a much more complicated definition. Finally, as an example of $\odot$, consider the case $r \geq q = k$. By direct computation using the symmetry of f,

$$f_q \odot (t) g_q(\sigma_1,\cdots,\sigma_{r-q})$$

$$= \frac{1}{(r-q)!}\frac{1}{q!}\int_0^t \cdots \int_0^t \sum_{S_{r-q}} f(s_1,\cdots,s_q,\sigma_{\tau(1)},\cdots,\sigma_{\tau(r-q)}) g(s_1,\cdots,s_r) \times ds_q \cdots ds_1$$

$$= \frac{1}{q!}\int_0^t \cdots \int_0^t f(s_1,\cdots,s_q,\sigma_1,\cdots,\sigma_{r-q}) g(s_1,\cdots,s_q) ds_q \cdots ds_1 .$$

The main theorem of this section is:

Theorem 2.1 Let $f \in \hat{L}^2([0,T]^r)$, $g \in \hat{L}^2([0,T]^q)$.

$$I_t^r(f)\ I_t^q(g) = \sum_{k=0}^{\min(r,q)} I_t^{r+q-2k}\left(\binom{r+q-2k}{r-k} f_k \odot (t) g_k\right) \quad . \qquad (2.14)$$

(2.14) shall be referred to as the multiplication formula. Our proof of (2.14) uses Ito's differentiation rule and induction, and it is fairly complicated in its details. Therefore, before embarking on the proof, we will set forth the relevant properties of $\odot$ in some preliminary lemmas. In what follows, f will always denote a function in $\hat{L}^2([0,T]^r)$, g a function in $\hat{L}^2([0,T]^q)$.

Lemma 2.4 For every $t \leq T$

$$f_k \odot (t) g_k \in L^2([0,T]^{r+q-2k})$$

In fact

$$|| f_k \odot (t) g_k ||^2 \leq c_{r,q,k}\ || f ||^2\ || g ||^2$$

where $c_{r,q,k}$ is independent of f and g.

Proof It suffices to prove the lemma for $\otimes$, instead of $\odot$, since P_{r+q-2k} is a bounded operator. Let $d\bar{\sigma} = d\sigma_1 \cdots d\sigma_{r+q-2k}$ $d\bar{s} = ds_1 \cdots ds_k$. We then have, using the Cauchy-Schwarz inequality

$$f_k \otimes (t) g_k = \int_{[0,T]^{r+q-2k}} d\bar{\sigma}\ \frac{1}{(k!)^2}$$

$$\times \left[\int_{[0,T]^k} d\bar{s} f(s_1,\cdots,s_k,\sigma_1,\cdots) g(s_1,\cdots,s_k,\cdots,\sigma_{r+q-2k}) \right]^2$$

$$\leq \frac{1}{(k!)^2} \int_{[0,T]^{r+q-2k}} d\bar{\sigma} \int_{[0,T]^k} d\bar{s}\ f^2(s_1,\cdots,s_k,\sigma_1,\cdots,\sigma_{r-k})$$

$$\times \int_{[0,T]^k} d\bar{s}\ g^2(s_1,\cdots,s_k,\sigma_{r-k+1},\cdots,\sigma_{r+q-2k})$$

$$= \frac{1}{(k!)^2} \|f\|^2 \|g\|^2 \quad .$$

Lemma 2.4 establishes that the kernels in the expansion 2.14 are square-integrable and hence that the multiple integrals are well-defined. The next lemma collects useful identifies and facts about $\odot$. Recall that the notation $f(s_1,\cdots,s_k,\cdots)$ indicates the section of f in which the first k variables are fixed at $s_1,\cdots,s_k$, respectively.

Lemma 2.5

(i) $f(\sigma_1,\cdots)_k \odot (\sigma_1) g(\sigma_1,\cdots)_k(\sigma_2,\cdots,\sigma_{r+q-2k-1}) \in \hat{L}^2([0,T]^{r+q-2k-1})$

(ii) $f_k \odot (t) g_k$

$$= f_k \odot (s) g_k - \int_\sigma^t f(s,\cdots)_{k-1} \odot (s) g(s,\cdots)_{k-1} ds \qquad (2.15)$$

(iii) For $k \geq 1$, $f_k \odot (t)\, g_k(\sigma_1,\cdots,\sigma_{r+q-2k})$

$$= \left[\frac{r-k}{r+q-2k} f(\sigma_1,\cdots)_k \odot (t) g_k + \frac{q-k}{r+q-2k} f_k \odot (t) g(\sigma_1,\cdots)_k \right]$$

$$(\sigma_2,\cdots,\sigma_{r+q-2k}) \qquad (2.16)$$

(iv) $f \odot (t) g(\sigma_1,\cdots,\sigma_{r+q}) = \left[\frac{r}{r-q} f(\sigma_1,\cdots) \odot (t) g \right.$

$$\left. + \frac{q}{r+q} f \odot (t) g(\sigma_1,\cdots)\right] (\sigma_2,\cdots,\sigma_{r+q-2k}). \qquad (2.17)$$

Proof

(i) follows by calculations similar to the proof of lemma 2.4, namely, one writes out the definition of the square norm and applies Cauchy-Schwarz. The details will not presented.

(ii) By direct calculation and definition, using the symmetry of f and g extensively.

$$f_k \odot (t) g_k$$

$$= P_{r+q-2k}\left[\frac{1}{k!} \int_0^t \cdots \int_0^t ds_1 \cdots ds_k f(s_1,\cdots,s_k,\cdots) g(s_1,\cdots s_k,\cdots)\right]$$

$$= P_{r+q-2k}\left[\int_0^t \int_0^{s_1} \cdots \int_0^{s_{k-1}} f(s_1,\cdots,s_k,\cdots) g(s_1,\cdots,s_k\cdots) ds_k \cdots ds_1\right]$$

$$= P_{r+q-2k}\left[\frac{1}{k!}\int_0^\sigma\int_0^\sigma\cdots\int_0^\sigma f(s_1,\cdot\cdot,s_k,\cdot\cdot)g(s_1,\cdot\cdot,s_k,\cdot\cdot)ds_k\cdot\cdot ds_1\right]$$

$$+ P_{r+q-2k}\left[\frac{1}{(k-1)!}\int_\sigma^t ds\int_0^{s_1}\cdots\int_0^{s_{k-1}} f(s_1,\cdot\cdot,s_k,\cdot\cdot)g(s_1,\cdot\cdot s_k,\cdot\cdot)ds_k\cdot\cdot ds_1\right]$$

$$= f_k \odot (\sigma) g_k + \int_\sigma^t ds\ f(s,\cdot\cdot)_{k-1} \odot (s) g(s,\cdot\cdot)_{k-1}$$

(iii) and (iv). The proofs of (iii) and (iv) are similar, (iv) being just a special case of (iii). We shall only present (iv) as it is simpler. Note first that, by definition,

$$\frac{r}{r+q}[f(\sigma_1,\cdot\cdot)\odot(t)g](\sigma_2,\cdot\cdot,\sigma_{r+q})$$

$$= \frac{r}{r+q}\ \frac{1}{(r+q-1)}\ \sum_{\tau\in S_{r+q-1}} f(\sigma_1,\sigma_{\tau(2)},\cdot\cdot,\sigma_{\tau(r)})$$

$$g(\sigma_{\tau(r+1)},\cdot\cdot,\sigma_{\tau(r+q)}) \qquad (2.18)$$

where $\pi \in S_{r+q-1}$ is interpreted as a permutation of $\{2,\cdot\cdot,r+q\}$. Now using the symmetry of f, (2.18) may be written as:

$$\frac{1}{(r+q)!}\sum_{j=1}^{r}\ \sum_{\tau\in S_{r+q-1}} f(\sigma_{\tau(2)},\cdot\cdot,\sigma_{\tau(j-2)},\overset{j^{th}\text{ position}}{\overset{\downarrow}{\sigma_1}},\sigma_{\tau(j-1)},\cdot\cdot,\sigma_{\tau(r)})\times$$

$$g(\sigma_{\tau(r-1)},\cdot\cdot,\sigma_{\tau(r+q)}) \qquad (2.19)$$

Using the expression analogous to (2.19) for $\frac{q}{r+q} f \odot (t) g(\sigma_1 \cdots)$

$$\{\frac{r}{r+q} f(\sigma_1,\cdots) \odot (t) g + \frac{q}{r+q} f \odot (t)\ g(\sigma_1,\cdots)\}\ (\sigma_2,\cdots,\sigma_{r+q})$$

$$= \frac{1}{(r+q)!}\Big[\sum_{j=1}^{r} \sum_{\pi\in S_{r+q-1}} f(\sigma_{\pi(2)},\cdots,\overset{j^{th}\text{ position}}{\overset{\downarrow}{\sigma_1}},\cdots,\sigma_{\pi(r)})$$

$$g(\sigma_{\pi(r+1)},\cdots,\sigma_{\pi(r+q)})$$

$$+ \sum_{j=1}^{q} \sum_{\pi\in S_{r+q-1}} f(\sigma_{\pi(2)},\cdots,\sigma_{\pi(r+1)})$$

$$g(\sigma_{\pi(r+2)},\cdots,\overset{j^{th}\text{ position}}{\overset{\downarrow}{\sigma_1}},\cdots,\sigma_{\pi(r+q)})\Big]$$

$$= \frac{1}{(r+q)!} \sum_{\pi\in S_{r+q}} f(\sigma_{\pi(1)},\cdots) g(\sigma_{\pi(r+1)},\cdots,\sigma_{\pi(r+q)})$$

$$= f \odot (t) g(\sigma_1,\cdots,\sigma_{r+q}) .$$

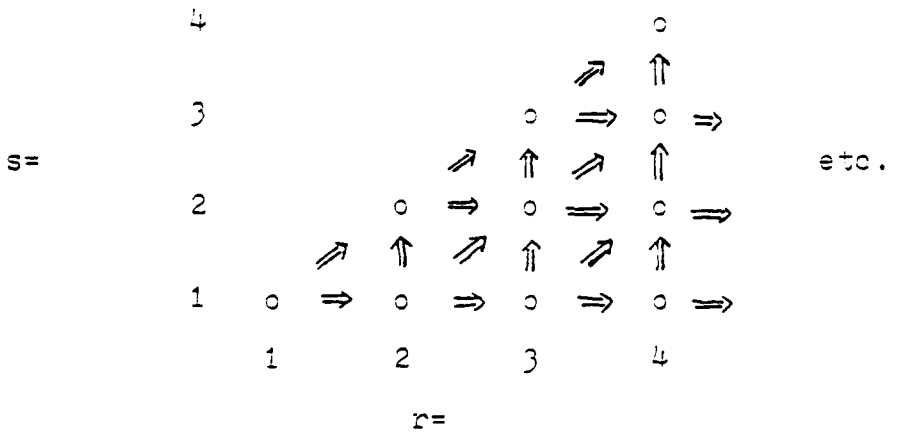
This is the desired result.

Proof of theorem 2.4. We use Ito's differentiation formula and the preceding lemmas to implement an induction argument that proceeds in two steps:

(a) Show (by induction) that (2.14) holds for orders $r = n$, $q = 1$, $\forall n$

(b) Assuming (2.14) for (r-1,q), (r,q-1) and (r-1,q-1), show that it holds for (r,q).

(a) and (b) then provide a consistent scheme of induction as shown by the following diagram demonstrating the paths of implication.



Step (a) For r = 1, q = 1. By Ito's differentiation rule

$$\int_0^t f(s)db(s) \int_0^t g(s)db(s) = \int_0^t \int_0^{s_1} [f(s_1)g(s_2)+f(s_2)g(s_1)]db(s_2)db(s_1)$$
$$+ \int_0^t f(s)g(s)ds \quad .$$

Suppose that the theorem is true for (r,q) = (n-1,1) and let $f \in \hat{L}^2([0,T]^n)$, $g \in \hat{L}^2([0,T])$. Applying Ito's differentiation rule again,

$$I_t^n(f)I_t^1(g) = \int_0^t g(s)\ I_s^n(f)db(s) + \int_0^t I_s^{n-1}(f(s,\cdot\cdot))I_s^1(g)db(s)$$

$$+ \int_0^t I_s^{n-1}(g(s)f(s,\cdot\cdot))ds \qquad (2.20)$$

By induction,

$$I_s^{n-1}(f(s,\cdot\cdot))I_s^1(g) = I_s^n(n[f(s,\cdot\cdot) \odot g]) + I_s^{n-2}(f(s,\cdot\cdot)_1 \odot (s)g_1).$$

Lemma 2.5(i) and lemma 2.1 justify interchanging integrations in the last term of (2.20):

$$\int_0^t I_s^{n-1}(g(s)f(s,\cdot\cdot))ds = I_t^{n-1}(\int_{s_1}^t g(u)f(u,s_1,\cdot\cdot,s_{n-1})du)\ .$$

Thus, by substitution in (2.20)

$$I_t^n(f)\ I_t^1(g) = \int_0^t \{I_s^n(g(s)f(\cdot\cdot)) + I_s^n(n[f(s,\cdot\cdot) \odot g])\}db(s)$$

$$+ \int_0^t I_s^{n-2}(f(s,\cdot\cdot)_1 \odot (s)g_1)db(s)$$

$$+ I_t^{n-1}(\int_{s_1}^t g(u)f(u,s_1,\cdot\cdot,s_{n-1})du)$$

$$= I_k^{n+1}\{g(\sigma_1)f(\sigma_2,\cdots,\sigma_n) + n[f(\sigma_1,\cdots)\odot(\sigma_1)](\sigma_2,\cdots,\sigma_n)\}$$

$$+ I_t^{n-1}\{f(\sigma_1,\cdots)_1\odot(\sigma_1)g_1(\sigma_2,\cdots,\sigma_n) +$$

$$+ \int_0^t g(s)f(s,\sigma_1,\cdots,\sigma_{n-1})ds\}.$$

And by lemma 2.5 (iii) and (iv) this becomes

$$I_t^{n+1}((n+1)f\odot g) + I_t^{n-1}(f_1\odot(t)g_1)\ ,$$

which completes the induction step of (a).

Step b Without loss of generality assume that $q \leq r$. The induction hypothesis is that theorem 2.4 is true for $(r-1,q)$, $(r,q-1)$, and $(r-1,q-1)$. Apply Ito's differentiation rule:

$$I_t^r(f)I_t^q(g) = \int_0^t I_s^q(g)I_s^{r-1}(f(s,\cdots))db(s)$$

$$+ \int_0^t I_s^{q-1}(g)I_s^r(f)db(s)$$

$$+ \int_0^t I_s^{r-1}(f(s,\cdots))I_s^{q-1}(g(s,\cdots))ds\ . \qquad (2.21)$$

By induction

$$I_s^q(g)I_s^{r-1}(f(s,\cdots) = \sum_{k=0}^{\min(q,r-1)} I_s^{r+q-1-2k}\left(\binom{r+q-1-2k}{r-1-k}f(s,\cdots)_k\odot g_k\right) \qquad (2.22)$$

$$I_s^r(f)I_s^{q-1}(g(s,\cdot\cdot)) = \sum_{k=0}^{\min(r,k)} I_s^{r+q-1-2k}\left(\binom{r+q-1-2k}{r-k} f_k \odot (s) g_k(s,\cdot\cdot)_k\right) \tag{2.23}$$

$$I_s^{r-1}(f(s,\cdot\cdot))I_s^{q-1}(g(s,\cdot\cdot))$$

$$= \sum_{k=1}^{\min(q-1,r-1)+1} I_s^{r+q-2k}\left(\binom{r+q-2k}{r-k} f(s,\cdot\cdot)_{k-1} \odot (s) g(s,\cdot\cdot)_{k-1}\right) \tag{2.24}$$

Now substitute (2.22) - (2.24) into (2.21), interchange dt and db(t) integrations where necessary, and collect like order terms. The result, after some nasty calculation, is, if $q < r$

$$\begin{aligned} I_t^r(f)I_t^q(g) &= I_t^{r+q}\{[\binom{r+q-1}{r-1}[f(s_1,\cdot\cdot) \odot g] + \binom{r+q-1}{r}[f \odot g(s_1,\cdot\cdot)]](s_2,\cdot\cdot,s_{r+q})\} \\ &+ \sum_{k=1}^{q-1} I_t^{r+q-2k}\{\binom{r+q-1-2k}{r-1-k}[f(s_1,\cdot\cdot)_k \odot (s_1) g(s_1,\cdot\cdot)_k](s_2,\cdot\cdot,s_{r+q-2k}) \\ &\quad + \binom{r+q-1-2k}{r-k}[f_k \odot (s_1) g(s_1,\cdot\cdot)_k](s_2,\cdot\cdot,s_{r+q-2k}) \\ &\quad + \binom{r+q-2k}{r-k}\int_{s_1}^{t} f(u,\cdot\cdot)_{k-1} \odot (u) g(u,\cdot\cdot)_{k-1}\,du\} \\ &+ I_t^{r-q}\{[f(s_1,\cdot\cdot)_q \odot (s_1) g_q](s_2,\cdot\cdot,s_{r-q}) \end{aligned} \tag{2.25}$$

$$+ \int_{s_1}^{t} f(u,\cdot\cdot)_{q-1} \odot (u) g(u,\cdot\cdot)_{q-1} du \Big\} .$$

Now examine the kernels of the last expression one-by-one. The first kernel equals

$$\binom{r+q}{r} \Big[\frac{r}{r+q} (f(s_1,\cdot\cdot) \odot g)(s_2,\cdot\cdot,s_{r+q})$$

$$+ \frac{q}{r+q} (f \odot g(s_1,\cdot\cdot))(s_2,\cdot\cdot,s_{r+q})$$

$$= \binom{r+q}{r} (f \odot g)(s_1,\cdot\cdot,s_{r+q}) . \qquad (2.27)$$

The last equality comes from lemma 2.5(iv). Likewise apply lemma 2.5(iii) and (iv) to the kernel of I_t^{r+q-2k} , $1 \le k \le q - 1$ The kernel of I_t^{r+q-2k} equals

$$\binom{r+q-2k}{r-k} \Big[\frac{r-k}{r+q-2k} (f(s_1,\cdot\cdot)_k \odot (s_1) g_k)(s_2,\cdot\cdot)$$

$$+ \frac{q-k}{r+q-2k} [f_k \odot (s_1) g(s_1,\cdot\cdot)_k] (s_2,\cdot\cdot)$$

$$+ \Big(\int_{s_1}^{t} f(u,\cdot\cdot)_{k-1} \odot (u) g(u,\cdot\cdot)_{k-1} du \Big)(s_2,\cdot\cdot) \Big]$$

$$= \binom{r+q-2k}{r-k} \Big[(f_k \odot (s_1) g_k)(s_1,\cdot\cdot,s_{r+q-2k})$$

$$+ \Big(\int_{s_1}^{t} f(u,\cdot\cdot)_{k-1} \odot (u) g(u,\cdot\cdot)_{k-1} du \Big)(s_2,\cdot\cdot$$

$$= \binom{r+q-2k}{r-k}[(f_k \odot (t)g_k)(s_1,\cdots,s_{r+q-2k})] \tag{2.28}$$

Finally, in the same way, the kernel for I_t^{r-q} is

$$(f_q \odot (t)g_q)(s_1,\cdots,s_{r-q}) \quad . \tag{2.29}$$

By combining (2.27) - (2.29) in (2.26) we complete the induction step (for $q < r$):

$$I_t^r(f)I_t^q(g) = \sum_{k=0}^{\min(r,q)} I_t^{r+q-2k}(\binom{r+q-2k}{r-k} f_k \odot (t)g_k) \quad .$$

The proof for $q = r$ is the same; we need only check that the lowest order contribution in (2.26) corresponding to $k = q$ is

$$f_q \odot (t)g_q = \int_0^t \int_0^{s_1} \cdots \int_0^{s_{q-1}} f(s_1,\cdots,s_q)g(s_1,\cdots,s_q)ds_q\cdots ds_1.$$

The multiplication formula relates directly to properties of Hermite polynomials, as one naturally suspects from the connection between homogeneous chaos and multiple Wiener integrals. In fact, letting $\{h_n(x)\}_{n=0}$ denote the Hermite polynomials defined in Appendix 1 and taking $\{\phi_n\}_{n=1}^{\infty}$ to be a complete orthonormal basis of $L^2([0,T]^r)$, recall from theorem 2.2 and theorem A.1.1 that

$$\{I_T^r(f) \mid f \in \hat{L}_2(LO,T]^r)\} = \overline{G}_r$$

$$= \overline{S_{pan}} \{ \prod_{i=1}^{n} h_{p_i}\left(\int_0^T \phi_{j_i}(\sigma)db(\sigma)\right) \mid p_1+\cdots+p_n=r, \quad j_i \text{ are}$$

$$\text{pairwise unequal} \} \quad . \tag{2.30}$$

Now, the $h_n(x)$ satisfy the identities

$$h_r(x)h_q(x) = \sum_{k=0}^{\min(r,q)} \sqrt{\binom{r}{k}\binom{q}{k}}\sqrt{\binom{r+q-2k}{r-k}}\, h_{r+q-2k}(x) \tag{2.31}$$

for $r,q \geq 0$ [see e.g., Magnus and Oberhettinger [26]). Comparing (2.31) to (2.14), one thus sees via (2.30) that the multiplication formula effectively generalizes the identities (2.31). There is a discrepancy between (2.31) and (2.14) in the factors multiplying the expansion terms, but this is due to the different normalizations involved in the definitions of h_n, I^r and $\odot$.

It is natural to ask whether theorem 2.4 can be proved using (2.31). However, this strategy appears exceedingly difficult to implement and I have not succeeded in doing so. Recently Hida [18] obtained a proof of the multiplication formula independently of myself. His proof effectively generalizes the techniques used to prove (2.31), but to do so he must invoke his advanced theory of generalized Brownian functionals. Our proof, though involved

computationally, proceeds via elementary methods.

The multiplication formula can also be used to generalize a fact about Gaussian random variables. Suppose Z is a normally distributed r.v. with mean 0 and variance σ^2. Then the well known identity

$$E\, Z^{2m} = (2m-1)\cdots 3\cdot\sigma^{2m}$$

expresses the higher order moments of Z in terms of the variance. Clearly it implies that if X_n is a sequence of mean 0 normal r.v.'s, $EX_n^2 \to 0$ as $n \to \infty$ iff $EX_n^{2m} \to 0$ as $n \to \infty$ for any integer $m \geq 1$, or, in other words, that mean square and $2m^{th}$ order convergence are equivalent for any given m. Now $I_t^r(f)$ is an element in the mean-square closure of r^{th} order polynomials of a Gaussian process, and hence itsmoment convergence properties are similar.

Theorem 2.5 For any r and k, there exists an $M_{r,k} < \infty$ such that

$$E\{(I_T^r(f))^{2k}\} \leq M_{r,k}\|f\|^{2k} \tag{2.32}$$

for all $f \in \hat{L}^2([0,T]^r)$.

Proof Assume that, for a given n,

$$(I_T^r(f))^n = \sum_{\ell=0}^{nr} I_T^\ell(k_\ell) \tag{2.33}$$

where

$$||k_\ell|| \leq M||f||^{2n}, \quad \ell \leq nr \tag{2.34}$$

Using the multiplication formula there exist kernels h_ℓ such that

$$(I_T^r(f))^{n+1} = \sum_{\ell=0}^{(n+1)r} I_T^\ell(h_\ell)$$

and, from lemma 2.4, there will exist an N such that

$$||h_\ell|| \leq N \sum_{j=0}^{nr} ||k_j||^2 \, ||f||^2 \leq N||f||^{2(n+1)}$$

for every ℓ. Thus (2.33) and (2.34) will hold for $n+1$ also, and, since the case, $n = 1$ is true, they will hold for all n by induction. But then

$$E[I_T^r(f)]^{2n} = \sum_{\ell=0}^{nr} E[I_T^\ell(k_\ell)]^2 \leq M' \sum_{\ell=0}^{nr} ||k_\ell||^2$$

$$\leq M'' \, ||f||^{2n}.$$

Of course M'' will depend on n and r.

<u>Corollary</u> Let $\{f_n\}_{n=1}^\infty$ and f be functions in $\hat{L}^2([0,T]^r)$. Then $||f_n - f|| \to 0$ as $n \to \infty$ iff

$$E[I_T^r(f_n)-I_T^r(f)]^{2k} \to 0 \qquad \text{as } n \to \infty$$

for any or all $k \geq 1$.

Remark Theorem 2.5 is not new. I. Segal [43] has derived (2.32) in the context of Gauss measures on Hilbert spaces. In fact, he obtains a universal constant c such that

$$E(I_T^r(f))^{2k} \leq 2^{2krc}\,|||f|||^{2k} \quad \text{for all } k.$$

Theorem 2.5 is also related to the Nelson [38] theory of hypercontractivity; see Mitter and Ocone [36]. Nevel [39] is a good reference for one version of Nelson's hypercontractivity theorem. McKean [32] and Wiener [45] also develop identities for expressions similar to $E[(I_T^r(f))^{2k}]$ in theorem 2.5. In these treatments, the interesting corollary above is not generally observed.

The next result is a variant of an easy and well-known identity first appearing in Ito [20], and stated here in (2.39). Our proof employs the multiplication formula.

Theorem 2.6 Let $\phi \in L^2(\Omega, \mathcal{F}_T^b, P)$ have the multiple integral expansion

$$\phi = \sum_{r=0}^{\infty} I_T^r(k_r) .$$

Then

$$k_r(\sigma_1,\cdots,\sigma_r) = \frac{\partial^r}{\partial\sigma_1\cdots\partial\sigma_r} E\,b(\sigma_1)\cdots b(\sigma_r) .$$

Proof For given $\sigma_0 \geq \sigma_1 \geq \cdots \geq \sigma_r$, let

$$\gamma^i(s) = 1_{[0,\sigma_i]}(s), \quad 0 \leq i \leq r$$

be the indicator function of $[0,\sigma_i]$. We will first show that

$$b(\sigma_1)\cdots b(\sigma_r) = I_1^r(\gamma^1\cdots\gamma^r) + \sum_{n=0}^{[\frac{r}{2}]} I_1^{r-2n}(h^n_{\sigma_1\cdots\sigma_r}) \qquad (2.35)$$

where

$$[\tfrac{r}{2}] \equiv \max\{n : n \leq \tfrac{r}{2}\}$$

and each kernel $h^n_{\sigma_1\cdots\sigma_r}$ has the form

$$h^n_{\sigma_1\cdots\sigma_r} = \sum_{i=1}^{N} g^i_{\sigma_1\cdots\sigma_r} \qquad (2.36)$$

in which the $g^i_{\sigma_1\cdots\sigma_r}$ functions each depend only on a proper subset of the indices $\sigma_1,\cdots,\sigma_r$. This fact will greatly simplify the task of calculating $\frac{\partial^r}{\partial\sigma_1\cdots\partial\sigma_r} E\,b(\sigma_1)\cdots b(\sigma_r)$ as will be seen in (2.39) below.

Since (2.35) is valid for $r = 1$, it can be proved for general r by induction. Thus, suppose (2.35) is true for r.

Then

$$b(\sigma_0)\cdots b(\sigma_r) = I_T^1(\gamma^0)I_T^r(\gamma^1\cdots\gamma^r)$$

$$+ \sum_{n=0}^{[r/2]} I_T(\gamma^0)I_T^{r-2n}(h^n_{\sigma_1\cdots\sigma_r})$$

Now expand all the products in this sum by the multiplication formula. It is easily seen that the kernels of any multiple integrals that arise from expanding the terms

$$I_T^1(\gamma^0)I_T^{r-2n}(h^n_{\sigma_1\cdots\sigma_r})$$

will be of the form (2.36). However, the first product is

$$I_T^1(\gamma^0)I_T^r(\gamma^1\cdots\gamma^r)$$

$$= I_T^{r+1}((r+1)\gamma^0 \odot (T)\gamma^1\cdots\gamma^r) + I_T^{r-1}((\gamma^0)_1 \odot (T)(\gamma^1\cdots\gamma^r)_1).$$

The first kernel is

$$\frac{1}{r+1} \sum_{j=0}^{r} \gamma^0(s_j)[\gamma^1(s_0)\cdots\gamma^j(s_{j-1})\gamma^{j+1}(s_{j+1})\cdots\gamma^r(s_r)]$$

for $s_0 \geq s_1\cdots\geq s_r$.

But note that if $j \geq 0$, $\gamma^0(s_j)\gamma^1(s_0)=\gamma^1(s_0)$ for $s_0 \geq s_j$, since $s_0 \geq s_1 \geq s_0$ implies $s_0 \geq s_j$. Thus the last expression may be written more simply as

$$\frac{1}{r+1}\gamma^0(s_0)\cdots\gamma^r(s_r) + \frac{1}{r+1}\sum_{j=1}^{r}\gamma^1(s_0)\cdots\gamma^j(s_{j-1})\gamma^{j+1}(s_{j+1})\cdots\gamma^r(s_r) \tag{2.37}$$

and only the first term depends on all the indices $\sigma_0,\sigma_1,\cdots,\sigma_r$. Likewise, the second kernel is seen to be

$$\int_0^{\sigma_1}\gamma^1(s)\gamma^2(s_1)\cdots\gamma^r(s_{r-1})\,ds\ . \tag{2.38}$$

The only kernel in (2.37) and (2.38) that depends on every index is

$$\frac{1}{r+1}\gamma^0(s_0)\cdots\gamma^r(s_r)\ .$$

Thus by substituting the results of these kernel computations into the expansion of $b(\sigma_0)\cdots b(\sigma_r)$, we find that (2.35) is true for $r+1$ as well. Thus (2.35) holds for all r by induction.

As a result of (2.35)

$$\frac{\partial^r}{\partial\sigma_1\cdots\partial\sigma_r}\,E_\phi\, b(\sigma_1)\cdots b(\sigma_r)$$

$$= \frac{\partial^r}{\partial\sigma_1\cdots\partial\sigma_r}\,E_\phi\, I_1^r(\gamma^1\cdots\gamma^r)$$

$$+ \frac{\partial^r}{\partial\sigma_1\cdots\partial\sigma_r}\sum_{n=0}^{[r/2]}E_\phi I_1^{r-2n}(h^n_{\sigma_1\cdots\sigma_r})$$

$$= \frac{\partial^r}{\partial\sigma_1\cdots\partial\sigma_r} \; E_0 \, I_T^r(\gamma^1\cdots\gamma^r)$$

$$= \frac{\partial r}{\partial\sigma_1\cdots\partial\sigma_r} \int_0^t \int_0^{s_1} \cdots \int_0^{s_{r-1}} k_r(s_1,\cdots,s_r) 1_{[0,\sigma_1]}(s_1)\cdots$$

$$1_{[0,\sigma_r]}(s_r) \quad ds_r\cdots ds_1 \qquad (2.39)$$

$$= k_r(\sigma_1,\cdots,\sigma_r) \; .$$

This completes the proof.

Finally, we note that theorem 2.4 extends easily to the case of multiple integrals with respect to observation semi-martingales. Under added assumptions, theorem 2.5 and its corollary extends also. Indeed, let $I^r_{T,y}(f)$ now denote integrals with respect to y and let P_0 be the measure w.r.t. which $y(\cdot)$ is Brownian.

Theorem 2.7 If $E_0 \left(\frac{dP}{dP_0}\right)^2 < \infty$, theorem 2.5 and its corollary hold for $I^r_{T,y}$.

Proof Use the Cauchy-Schwarz inequality to derive

$$E[I^r_{T,y}(f)]^{2k} = E_0\left\{\frac{dP}{dP_0} \, I^r_{T,y}(f)\right\} \le E_0^{1/2} \left[\frac{dP}{dP_0}\right]^2 E_0^{1/2} \left[I^r_{T,y}(f)\right]^{4k}$$

$$\le E_0^{1/2} \left(\frac{dP}{dP_0}\right)^2 M_{r,k} \|f\|^{2k}.$$

Remark

$$E_0 \left(\frac{dP}{dP_0}\right)^2 = E \exp \int_0^T x^2(s)\, ds.$$ See, e.g., Wong [46].

CHAPTER 3 ESTIMATION OF NONLINEAR FUNCTIONS OF GAUSSIAN PROCESSES

In this chapter, we begin the application of multiple integral expansions by treating a filtering problem considered by S. Marcus in his thesis [29], (see also Marcus and Willsky [31]). Marcus succeeding in constructing a class of filtering models which are interesting for systems applications and for which optimal filters can be finite in dimensionally and recursively computed, (indeed, a rare and happy event!). Roughly speaking, these models pass the outputs x(t) of linear stochastic systems through polynomial nonlinearities and seek to estimate the result based on linear observations of x(t) in white noise.

Marcus's original proofs accordingly rely upon linear filtering theory and Gaussian moment identities, and so his techniques never really leave the realm of Gaussian process theory. One naturally suspects that the proper framework for his problem is homogeneous chaos theory, the theory of polynomials of Gaussian processes. In what follows, we will show this suspicion to be well founded by developing a direct proof of Marcus's results with multiple Wiener integral techniques. We feel this proof explains in the clearest manner why finite dimensional filters occur in this problem and how the filter statistics arise. The results to to be discussed here have appeared in Marcus, Mitter, Ocone [30],

where an extension by Marcus to the discrete time case is also presented. It should be noted, as an aside, that solving the discrete time case requires using polynomials of the innovations, (see Marcus, et. al. [30]). This feature, which does not occur for continuous time, is explained by homogeneous chaos theory. We will not comment on the discrete case any further. Finally, Hida and Kallianpur [19] solve the related problem of predicting polynomials of a Gaussian process using noiseless observations, and they also use multiple integral techniques. By contrast, the results here deal with the case of noisy observations.

A brief exposition of the homogeneous chaos theory relevant to this chapter is presented in Appendix 1.

The problem may be stated as follows. Consider the linear system:

$$dx(t) = F(t)x(t)dt + G(t)dw(t) \qquad x(0)=x_0 \tag{3.1a}$$

$$dy(t) = H(t)x(t)dt + d\eta(t) \qquad y(0)=0 \tag{3.1b}$$

In (3.1): $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^n$; F, G, and H are piecewise continuous, bounded matrix valued functions; $w(t)$ and $\eta(t)$ are independent, vector Brownian motions; and x_0 is a Gaussian random variable independent of both $w(\cdot)$ and $\eta(\cdot)$. We consider $x(t)$ to be the signal process, $y(t)$ the observation process, and we are interested in calculating the filter $\hat{f}_t = E[f_t(x(s),s\leq t) \mid \mathcal{Y}_t]$ for functionals

f_t which are finite Volterra series with separable kernels in the elements of $x(t)$, (see definition 3.1).

In 2.1, we gave a construction for realizing a multiple stochastic integral with a separable kernel as the output of a stochastic differential system. By the same construction, we can produce f_t, i.e., there exist matrices $A_\ell(t)$, $\ell=1,\cdots,m$ and a vector $c(t)$ such that

$$f_t = c^T(t)z(t) \tag{3.2}$$

where

$$\frac{dz}{dt} = \left[\sum_{\ell=1}^{m} A_\ell(t)x_\ell(t)\right] \quad z(t) \quad z(0)=0 \tag{3.3}$$

Figure 1 illustrates the situation:

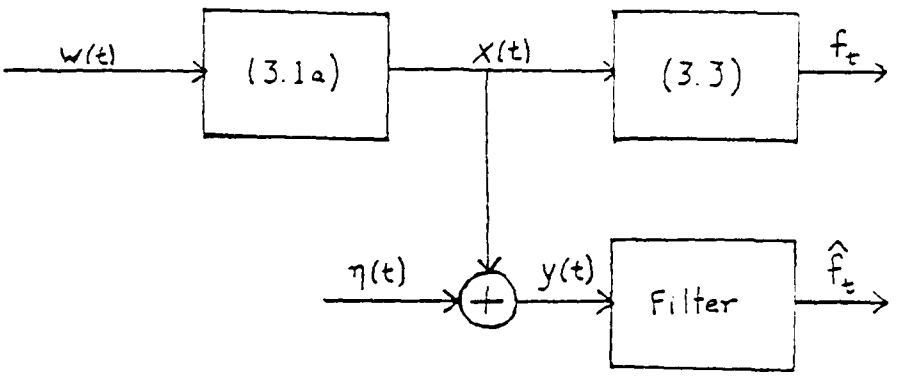


Fig. 1

The following definitions are convenient:

Definition 3.1

$$\Lambda_p \equiv \{ \sum_{i=1}^{m} \int_0^t \cdots \int_0^{\sigma_{p-1}} \gamma_0^i(t)\cdots\gamma_p^i(\sigma_p)$$

$$x_{k_{i,1}}(\sigma_1)\cdots x_{k_{i,p}}(\sigma_p)d\sigma_p\cdots d\sigma_1 \mid M<\infty, \gamma_\ell^k(\cdot) \in L^2_{loc}(\mathbb{R})\}$$

(where $\Lambda_0 \equiv \{\gamma_0(t) \mid \gamma_0 \in L^2_{loc}(\mathbb{R})\}$)

Definition 3.2 The filter $\hat{f}_t = E\{f_t(x(s), s\leq t) \mid \mathcal{F}_t^y\}$ is finite-dimensionally computable (FDC), if it can be computed from the output of a finite dimensional stochastic differential equation driven by $y(t)$.

Marcus [29] proved:

Theorem 3.1 For any p, if $f_t \in \Lambda_p$, $\hat{f}_t$ is finite-dimensionally computable.

Remark The theorem remains true if $dy(t) = H(t)x(t)dt + R(t)dw(t)$ where $R(t)>0$ and is deterministic. The proof is a trivial

adaption of the one to follow.

To carry out the proof, we need some properties of the state estimator $\hat{x}(t) = E[x(t)|F_t^y]$.

Lemma 3.1 (Kalman-Bucy)

i)

$$d\hat{x}(t) = F(t)\hat{x}(t)dt + P(t)H^T(t)dv(t) , \quad \hat{x}(0) = x_0$$

where

$$v(t) \equiv y(t) - \int_0^t H(s)\hat{x}(s)ds$$

and $P(t) \equiv E[(x(t)-\hat{x}(t))\,(x(t)-\hat{x}(t))^T]$ satisfies

$$\dot{P}(t) = F(t)P(t) + P(t)F^T(t) + G(t) - P(t)H^T(t)H(t)P(t)$$

$$P(0) = \mathrm{cov}(x_0)$$

ii) $v(t)$ is a Brownian motion and $F_t^v = F_t^y$ (up to sets of measure zero) for all t.

Proof See, for instance, M.H.A. Davis [9,10]. The process $v(t)$ defined in lemma 3.1 is called the [illegible] process, and

it is the key to our proof, because it is a Brownian motion that captures the information in $y(\cdot)$. That is, if f is a random variable such that $Ef^2<\infty$, then, by lemma 3.1 ii) $E\{f|F_t^y\} = E\{f|F_t^\nu\}$ and, hence, by the homogeneous chaos expansion, we can write

$$E\{f|F_t^y\} = k_o(t) + \sum_{i=1}^{n} \int_0^t k_1^i(t,\sigma)d\nu_i(\sigma)$$

$$+ \sum_{i,J=1}^{n} \int_0^t \int_0^{\sigma_1} k_2^{i,j}(t,\sigma_1,\sigma_2)d\nu_i(\sigma_2)d\nu_j(\sigma_1)$$

$$+ \cdots \qquad (3.4)$$

By using the innovations process, we thus achieve an orthogonal decomposition of any filter. But if $f \in \Lambda_p$, we can go much further.

Lemma 3.2 If $f_t \in \Lambda_p$, the expansion (3.4) truncates at order p:

$$\hat{f}_t = k_o(t) + \sum_{i=1}^{n} \int_0^t k_1^i(t,\sigma)\, d\nu_i(\sigma) + \cdots$$

$$+ \sum_{\ell_1\cdots\ell_p=1}^{n} \int_0^t \cdots \int_0^{\sigma_{p-1}} k_p^{\ell_1\cdots\ell_p}(t,\sigma_1,\cdots\sigma_p)$$

$$d\nu_{\ell_p}(\sigma_p)\cdots d\nu_{\ell_1}(\sigma_1) \qquad (3.5)$$

Proof: From the definitions of $x(t)$ and $\Lambda_p, Ef_t^2<\infty$ and thus (3.4) is valid. Now consider the process $z^T(t) = (x^T(t),\hat{x}^T(t),\nu^T(t))$. By (3.1) and Lemma 1 i)

$$dz(t) = \begin{bmatrix} F(t) & 0 & 0 \\ H(t) & F(t)-P(t)H^T(t)H(t) & 0 \\ H(t) & -H(t) & 0 \end{bmatrix} z(t)dt$$

$$+ \begin{bmatrix} G(t) & 0 \\ 0 & P(t)H^T(t) \\ 0 & I \end{bmatrix} \begin{bmatrix} dw(t) \\ d\eta(t) \end{bmatrix}$$

$$z(0)^T = (x_0^T, x_0^T, 0)$$

Thus $z(t)$ is generated from a linear system with Gaussian input and Gaussian initial value and hence is Gaussian. We conclude that $(x^T(t), \nu^T(t))$ is Gaussian also, and hence that we may apply the homogeneous chaos construction to it. In Appendix 1, this construction is developed for scalar processes, but it easy to see that the entire theory remains valid if we replace H of Appendix 1 by

$$H \equiv \overline{S}_{pan}\{x_i(s), \nu_j(s) \mid 1 \le i \le m, 1 \le j \le n, s \le t\}$$

Letting $A = \sigma\{x_i(s), v_j(s) \mid 1 \le i \le m, 1 \le j \le n, s \le t\}$, we can build from H the homogeneous chaos decomposition

$$L^2(\Omega, A, P) = \bigoplus_{\ell=0}^{\infty} \overline{G}_\ell$$

However, we can also perform the homogeneous chaos decomposition on the process $v(\cdot)$, that is, if $\overline{G_\ell(v)}$ denotes the ℓ^{th} homogeneous chaos of $v(s)$, $s \le t$,

$$L^2(\Omega, F_t^v, P) = \bigoplus_{\ell=0}^{\infty} \overline{G_\ell(v)}$$

Evidently, $\overline{G_\ell(v)} \subset \overline{G}_\ell$ for each ℓ and $f_t \in \bigoplus_{\ell=0}^{p} \overline{G}_\ell$ since f_t is a polynomial in $x(\cdot)$ of order p. Since $\overline{G_k(v)} \subset \overline{G_k}$ and $G_k \perp \bigoplus_{\ell=0}^{p} \overline{G}_\ell$ for $k > p$,

$$f_t \perp \overline{G}_\ell(v), \quad \ell > p.$$

But

$$E\{f_t \mid F_t^v\} = \text{projection of } f_t \text{ onto } \bigoplus_{\ell=0}^{\infty} \overline{G_\ell(v)}$$

Thus $E\{f_t \mid F_t^v\} \in \bigoplus_{\ell=0}^{p} \overline{G_\ell(v)}$ as desired.

We shall also need

Lemma 3.3 Let $z = [z_1,\cdots,z_k]$ be a jointly Gaussian random vector.

(i) $$E[z_1\cdots z_k] = Ez_1\, Ez_2\cdots z_k + \sum_{j=2}^{k} \operatorname{cov}[z_1,z_j]\, E[\prod_{\ell\neq j} z_\ell]$$

(ii) $$E[z_1\cdots z_k] = Ez_1\cdots Ez_k + \sum \operatorname{cov}(z_{j_1},z_{j_2})Ez_{j_3}\cdots Ez_{j_k}$$

$$+ \sum \operatorname{cov}(z_{j_1},z_{j_2})\operatorname{cov}(z_{j_3},z_{j_4})Ez_{j_5}\cdots Ez_{j_k}$$

$$+ \cdots$$

where the sums are taken over all combinations of pairs in $\{1\cdots k\}$

Proof These types of results are well-known; a good reference is K.S. Miller [34]. The particular form used here is that given in Marcus and Willsky [31].

Recall again that when a multiple Wiener integral has a separable kernel it is finite-dimensionally computable. Therefore, by lemma 2, $\hat{f}_t$ will be FDC if the kernels $k_r^{\ell_1\cdots\ell_r}(t,\sigma_1,\cdots,\sigma_r)$, $r \leq p$, of (3.5) are separable. Proving separability is thus the strategy of the proof of theorem 1. We shall need one more lemma that is a standard fact about linear stochastic differential equations.

Lemma 3.4 Let $Z(t)$ be the $\mathbb{R}^n$ valued solution of a linear system with Brownian inputs.

$$dZ(t) = A(t)Z(t)\,dt + B(t)dW(t) .$$

Here $W(t)$ is a vector Brownian motion and $A(t)$ and $B(t)$ are piecewise continuous matrices of appropriate dimension. Then

$$E[Z(t)-EZ(t)][Z(s)-EZ(s)]^T$$

$$= 1_{\{t<s\}}\, X_1(t,s) + 1_{\{t\geq s\}}\, X_2(t,s)$$

where $1_{\{t<s\}}$ denotes the indicator function of $\{t<s\}$ and $X_1(t,s)$ and $X_2(t,s)$ are matrices of separable functions.

Proof Let $\Phi(t,s)$ denote the state transition matrix of $A(t)$, that is $\frac{d}{dt}\Phi(t,s) = A(t)\Phi(t,s)$, $\Phi(s,s) = I$. Let

$$K(s) = \mathrm{Cov}[Z(s)] .$$

One easily calculates from $Z(t) = \Phi(t,s)Z(s) + \int_s^t \Phi(t,u)B(u)dw(u)$ that

$$E[Z(t)-EZ(t)][Z(s)-EZ(s)]^T = \begin{cases} \Phi(t,0)\Phi^{-1}(s,0)K(s) & s\leq t \\ K(s)[\Phi(s,0)\Phi^{-1}(t,0)]^T & s>t \end{cases}$$

Proof of Theorem 1* We can assume that f_t has the form

$$f_t = \int_0^t \cdots \int_0^{\sigma_{p-1}} \gamma_0(t)\gamma_1(\sigma_1)\cdots\gamma_p(\sigma_p)x_{k_1}(\sigma_1)\cdots x_{k_p}(\sigma_p)d\sigma_p\cdots d\sigma_1 \quad (3.6)$$

Then f_t has a finite expansion as in (3.5). The proof will consist of showing that the kernels $k_r^{\ell_1\cdots\ell_r}(t,\sigma_1,\cdots,\sigma_r)$, $r\leq p$, of the expansion (3.5) are separable. The idea is to use theorem 2.6 to express $k_r^{\ell_1\cdots\ell_r}$ in terms of f_t and ν_i:

$$k_r^{\ell_1\cdots\ell_r}(t,\sigma_1,\ldots,\sigma_r) = \frac{\partial^r}{\partial\sigma_1\cdots\partial\sigma_r} E[\hat{f}_t\nu_{\ell_1}(\sigma_1)\cdots\nu_{\ell_r}(\sigma_r)]$$

$$= \frac{\partial^r}{\partial\sigma_1\cdots\partial\sigma_r} E[f_t\nu_{\ell_1}(\sigma_1)\cdots\nu_{\ell_r}(\sigma_r)] \quad (3.7)$$

for $t \geq \sigma_1 > \sigma_2 > \ldots > \sigma_r$. The second equality in (3.7) is derived from the fact that $\nu(\sigma)$ is F_t^y-measurable for $\sigma \leq t$. When the expression (3.6) for f_t is used in (3.7), the result is:

$$k_r^{\ell_1\cdots\ell_r}(t,\sigma_1,\ldots,\sigma_r)$$

$$= \int_0^t \cdots \int_0^{s_{p-1}} \gamma_0(t)\cdots\gamma_p(s_p)\frac{\partial^r}{\partial\sigma_1\cdots\partial\sigma_r} E[x_{r_1}(s_1)\cdots x_{r_p}(s_p)\nu_{\ell_1}(\sigma_1)\cdots\nu_{\ell_r}(\sigma_p)]$$

$$ds_p\cdots ds_1 \quad (3.8)$$

*The idea of this proof is due to S. Marcus. It appears in Marcus, Mitter, and Ocone [30] but with some errors. The errors are corrected here.

Since $(x(t),\nu(t))$ is a Gaussian process, lemma 3,3ii) may be applied to the expectation term in the integrand of (3,8). The result is that this term may be written as a sum of products of the following terms

$$Ex_{k_i}(s_i) \ , \ \mathrm{cov}[x_{k_i}(s_i),x_{k_j}(s_j)],$$
$$\frac{\partial}{\partial\sigma_\ell}\mathrm{cov}[x_{k_i}(s_i),\nu_{\ell_j}(\sigma_j),$$
$$\frac{\partial^2}{\partial\sigma_i\partial\sigma_j}\mathrm{cov}[\nu_{\ell_i}(\sigma_i),\nu_{\ell_j}(\sigma_j)],$$

and

$$\frac{\partial}{\partial\sigma_i}E\nu_{\ell_i}(\sigma_i) \ .$$

Because $\nu(\cdot)$ is a Brownian motion, the last two terms are identically zero. The first two terms are separable functions of $(s_1,\ldots,s_p)$ on the range of integration $s_1 \geq s_2 \geq \ldots \geq s_p$; to see this for $\mathrm{cov}[x_{k_i}(s_i),x_{k_j}(s_j)]$ use lemma 3.4 and the fact that $s_i \geq s_j$ if $i < j$ on the range $s_1 \geq s_2 \geq \ldots \geq s_p$. The remaining term is

$$\frac{\partial}{\partial\sigma}\mathrm{cov}[x_k(s),\nu_\ell(\sigma)] = \frac{\partial}{\partial\sigma}\mathrm{cov}[x_k(s), \int_0^\sigma [H(s)(x(s)-\hat{x}(s))]_\ell ds$$
$$+ \eta_\ell(\sigma)] \qquad (3.9)$$

(Note: In (3.9) some of the subscripts have been dropped for simplicity of notation.) Since the observation noise $\eta(\cdot)$ is independent of the signal $x(\cdot)$, $\mathrm{cov}[x_p(s),\eta_\ell(\sigma)] \equiv 0$. Thus,

$$(3.9) = \mathrm{cov}[x_p(s), [H(\sigma)(x(\sigma)-\hat{x}(\sigma))]_\ell] \qquad (3.10)$$

But $Z^T(t) = (x^T(t),\hat{x}^T(t))$ is the solution of a linear stochastic

system with Brownian inputs. This system may be explicitly constructed using the Kalman-Bucy result in lemma 3.1. Thus, using lemma 3.4,

$$\begin{aligned}\operatorname{cov}[x_r(s),[H(\sigma)(x(\sigma)-x(\sigma))]_\ell] &= 1_{\{s \le \sigma\}}\psi_1(s,\sigma) + 1_{\{s > \sigma\}}\psi_2(s,\sigma) \\ &= 1_{\{s \le \sigma\}}[\psi_1(s,\sigma)-\psi_2(s,\sigma)] \\ &\quad + \psi_2(s,\sigma) \qquad (3.11)\end{aligned}$$

where ψ_1 and ψ_2 are separable functions.

The end result of all these remarks is that (3.3) may be written as a sum of expressions of the form

$$\int_0^t \cdots \int_0^{s_{p-1}} \gamma_0(t)\ldots\gamma_p(s_p)\alpha_1(s_1)\ldots\alpha_p(s_p)\beta(\sigma_1)\ldots\beta(\sigma_r) \times 1_{\{s_{j_1} \le \sigma_{i_1}\}}\ldots 1_{\{s_{j_q} \le \sigma_{i_q}\}}\, ds_p\ldots ds_1 \qquad (3.12)$$

such that $q \le r$

$$\{j_1,\ldots,j_p\} \subset \{1,\ldots,p\}$$

and

$$\{i_1,\ldots,i_p\} \subset \{1,\ldots,r\} .$$

To complete the proof it is only necessary to show that (3.12) is separable as a function of $(t,\sigma_1,\ldots,\sigma_r)$. However, by appropriately adjusting limits of integration in (3.12) we can write (3.12) in turn as a sum of terms of the sort

$$\int_{a_1}^{b_1} \cdots \int_{a_p}^{b_p} \gamma_0(t)\ldots\gamma_p(s_p)\alpha_1(s_1)\ldots\beta_r(\sigma_r)\,ds_p\ldots ds_1 \qquad (3.13)$$

Since $(x(t),\nu(t))$ is a Gaussian process, lemma 3.3ii) may be applied to the expectation term in the integrand of (3.8). The result is that this term may be written as a sum of products of the following terms

$$Ex_{k_i}(s_i)\ ,\ \mathrm{cov}[x_{k_i}(s_i),x_{k_j}(s_j)],$$
$$\frac{\partial}{\partial\sigma_j}\mathrm{cov}[x_{k_i}(s_i),\nu_{\ell_j}(\sigma_j)],$$
$$\frac{\partial^2}{\partial\sigma_i\partial\sigma_j}\mathrm{cov}[\nu_{\ell_i}(\sigma_i),\nu_{\ell_j}(\sigma_j)],$$

and

$$\frac{\partial}{\partial\sigma_i}E\nu_{\ell_i}(\sigma_i)\ .$$

Because $\nu(\cdot)$ is a Brownian motion, the last two terms are identically zero. The first two terms are separable functions of $(s_1,\ldots,s_p)$ on the range of integration $s_1 \geq s_2 \geq \ldots \geq s_p$; to see this for $\mathrm{cov}[x_{k_i}(s_i),x_{k_j}(s_j)]$ use lemma 3.4 and the fact that $s_i \geq s_j$ if $i < j$ on the range $s_1 \geq s_2 \geq \ldots \geq s_p$. The remaining term is

$$\frac{\partial}{\partial\sigma}\mathrm{cov}[x_k(s),\nu_\ell(\sigma)] = \frac{\partial}{\partial\sigma}\mathrm{cov}[x_k(s),\int_0^\sigma [H(s)(x(s)-\hat{x}(s))]_\ell ds + \eta_\ell(\sigma)] \qquad (3.9)$$

(Note: In (3.9) some of the subscripts have been dropped for simplicity of notation.) Since the observation noise $\eta(\cdot)$ is independent of the signal $x(\cdot)$, $\mathrm{cov}[x_p(s),\eta_\ell(\sigma)] \equiv 0$. Thus,

$$(3.9) = \mathrm{cov}[x_p(s),[H(\sigma)(x(\sigma)-\hat{x}(\sigma))]_\ell] \qquad (3.10)$$

But $Z^T(t) = (x^T(t),\hat{x}^T(t))$ is the solution of a linear stochastic

such that, for each i, $a_i \in \{0, \sigma_1, \ldots, \sigma_r\}$, $b_i \in \{s_{i-1}, \sigma_1, \ldots, \sigma_p\}$, $a_1 \geq a_2 \geq \ldots \geq a_p$, $s_i \geq a_{i+1}, i \leq p$. Note that (3.13) contains no indicator functions $1_{\{s \leq \sigma\}}$. Using the identity

$$\int_a^b g(s)ds = \int_0^b g(s)ds - \int_0^a g(s)ds$$

to write single integrals as separable functions of their upper and lower limits, it is seen that (3.13) will be a separable function of $(t, \sigma_1 \ldots \sigma_r)$. Thus (3.12) and (3.8) will also be separable, since they are ultimately sums of terms like (3.13). It is worthwhile illustrating the last argument with an example.

Consider

$$\int_0^t \int_0^{s_1} \alpha_1(s_1)\alpha_2(s_2)\beta_1(\sigma_1)\beta_2(\sigma_2)\, 1_{\{s_1 \leq \sigma_1\}} 1_{\{s_2 \leq \sigma_2\}} ds_2 ds_1$$

By straightforward calculation, this equals

$$\beta_1(\sigma_1)\beta_2(\sigma_2)\left\{\int_0^{\sigma_2}\int_0^{s_1} \alpha_1(s_1)\alpha_2(s_2)ds_2ds_1 + \int_{\sigma_2}^{\sigma_1}\int_0^{s_1} \alpha_1(s_1)\alpha_2(s_2)ds_2ds_1\right\}$$

$$= \beta_1(\sigma_1)\beta_2(\sigma_2)\left\{\int_0^{\sigma_2}\int_0^{s_1} \alpha_1(s_1)\alpha_2(s_2)ds_2ds_1 + \right.$$

$$\left. \int_0^{\sigma_2} \alpha_2(s_2)ds_2 \left[\int_0^{\sigma_1} \alpha_1(s_1)ds_1 - \int_0^{\sigma_2} \alpha_1(s_1)ds_1\right]\right.$$

which is separable.

CHAPTER 4 MULTIPLE STOCHASTIC INTEGRALS AND NONLINEAR FILTERING

This chapter applies the viewpoint of multiple integral expansions to the general filtering problem stated in the introduction. First, the Kallianpur-Striebel formula is used in 4.1 to derive a representation for the optimal filter as a ratio of two multiple integral series. The integrals in this representation are formed with respect to the observation process and have kernels that depend only upon the unconditioned distribution of the signal process and that hence may be computed offline, prior to receiving any observations. Secondly, we discuss the class of suboptimal filters consisting of a multiple integral expansion truncated after a finite number of terms. By combining the exact filter expansions, the multiplication formula, and change of measure, we derive kernel equations for the kernels of the best r^{th} order filter of this class. We then treat the cases $r = 1$ and $r = 2$ as examples and, using the same techniques, rederive the Kalman filter.

The filter expansion presented in 4.1 resembles formulae obtained by Eterno [11] in his thesis. Eterno built filter approximations by expanding the unnormalized conditional density in moment or cumulant power series, and his expressions, when appropriately evaluated, have multiple integral interpretations. Our expansion, which can also be applied to the conditional density, is constructed along different lines and we apply it to a different

class of suboptimal filter designs.

It is worthwhile empnasizing that the stochastic integrals employed in this section are taken with respect to the observation process, and not, as in Chapter 3, with respect to the innovations process. At first, integration against innovations appears to be an attractive alternative, because the innovations are Brownian and so allow one to exploit the homogeneous chaos theory, theorem 2.6 for computing kernels, and etc. in approximating filters. However, in constrast with Marcus' problem, the innovations are not easily calculated for they require optimally estimating the signal $h(x_t)$ (see the introduction), a problem of equivalent difficulty to the original one of estimating an arbitrary functional $f(x_s, s \leq t)$. Integrals against the observation process, on the other hand, are more readily computable, but less easy to handle, since $y(\cdot)$ is not in general Brownian, much less even Gaussian. $y(\cdot)$-based integrals of different orders are, for example, not orthogonal, making it difficult to project random variables on finite order sums of integrals. The technique introduced below to derive kernel equations for best finite order estimates addresses precisely this complication and provides tools for exploring the probabilistic structure of multiple integrals of $y(t)$ in more detail. Thus integrals of $y(t)$ can be analyzed and are more satisfactory for applications than integrals of the innovations process.

4.1 Filter Expansions

To fix notation, let us restate the general filtering problem and the Kallianpur-Striebel formula for the optimal estimate. $\{y(t)|0 \le t \le T\}$, $\{x(t)|0 \le t \le T\}$ and $\{w(t)|0 \le t \le T\}$ are scalar valued processes on the probability space (Ω,F,P), and $h(s,x)$ is a real-valued (Borel) function such that

$$\begin{aligned} y(t) &= \int_0^t h(s,x(s))ds + w(t) \qquad t \le T, \\ E\{&\int_0^T h^2(s,x(s))ds\} < \infty, \end{aligned} \tag{4.1}$$

and $w(\cdot)$ is standard Brownian, independent of $x(.)$.

Let $G_t = F_t^x \vee F_t^y$ and define P_0 by

$$\frac{dP_0}{dP} = \exp[-\int_0^T h(s,x(s))dw(s) - \frac{1}{2}\int_0^T h^2(s,x(s))ds] .$$

Recall that P_0 is a probability measure w.r.t. which $y(\cdot)$ and $x(\cdot)$ are independent, $x(\cdot)$ has the same law as under P, and $(y(t),G_t)_{t\le T}$ is a Wiener process. By this last statement, we mean that w.r.t. P_0, $y(t)$ is a continuous G_t-martingale such that $E_0\{[y(t)-y(s)]^2|G_s\} = t - s$: in general, $F_t^y \subsetneq G_t$. Let $f_t(x_s,s\le t)$ be a finite variance non-anticipative functional of $x(\cdot)$. For simplicity, we shall in the sequel always denote $f_t(x_s,s\le t)$ by $f(t)$.

and likewise $h(s,x(s))$ by $h(s)$. Then the Kallianpur-Striebel formula states

$$\hat{f}_t \equiv E\{f(t)|\mathcal{F}_t^y\} = \frac{E_0\{f(t)\exp[\int_0^t h(s)dy(s) - \frac{1}{2}\int_0^t h^2(s)ds]|\mathcal{F}_t^y\}}{E_0\{\exp[\int_0^t h(s)dy(s) - \frac{1}{2}\int_0^t h^2(s)ds]\}}. \tag{4.2}$$

Because of its importance, we single out the exponential term in (4.2) with the notation

$$L_t \equiv \exp[\int_0^t h(s)dy(s) - \frac{1}{2}\int_0^t h^2(s)ds] \quad .$$

It is well known, (Wong [46]), that $L_T = \frac{dP}{dP_0}$, L_t is a $\mathcal{G}_t$-martingale on $(\Omega,\mathcal{F},P_0)$ and

$$dL_t = h(t)L_t dy(t), \quad L_0 = 1 \tag{4.3}$$

(4.3) is the crucial relation for what follows.

In order to state the main theorem, it is convenient to introduce the functions

$$\begin{aligned} \gamma_n(t,s_1,\cdots,s_n) &= E\{f(t)h(s_1)\cdots h(s_n)\} \qquad n \geq 0 \\ k_n(s_1,\cdots,s_n) &= E\{h(s_1)\cdots h(s_n)\} \qquad n \geq 1 \ . \end{aligned} \tag{4.4}$$

Note that in (4.4) the expectation operates on random variables which depend only on the $x(\cdot)$ process, whose law is invariant under the change from P to P_0. Hence, we can also write

$$\begin{aligned} \gamma_n(t,s_1,\cdots,s_n) &= E_0\{f(t)h(s_1)\cdots h(s_n)\} \\ k_n(s_1,\cdots,s_n) &= E_0\{h(s_1)\cdots h(s_n)\} \ . \end{aligned} \tag{4.5}$$

Another useful process is

$$L_t^{(r)} = \int_0^t \cdots \int_0^{s_r} h(s_1)\cdots h(s_r) L_{s_{r+1}} dy(s_{r+1}) \cdots dy(s_1) .$$

The existence of $L_t^{(r)}$, a multiple stochastic integral with random coefficients, will be justified shortly. We now state multiple integral expansions for $\hat{f}_t$.

Theorem 4.1

(i) If $E[\int_0^T h^2(\sigma)d\sigma]^r < \infty$ and $E[f^2(t)(\int_0^T h^2(\sigma)d\sigma)^r] < \infty$,

$$\hat{f}_t = \frac{\lambda_0(t) + \sum_{n=1}^{r} I_t^{(r)}(\lambda_n) + E_0\{f(t)L_t^{(r)} | \mathcal{F}_t^y\}}{1 + \sum_{n=1}^{r} I_t^{(r)}(k_n) + E_0\{L_t^{(r)} | \mathcal{F}_t^y\}} . \tag{4.6}$$

(ii) If $E[\exp \int_0^T h^2(s)ds] < \infty$ and $E[f^2(t)\exp \int_0^T h^2(s)ds] < \infty$

$$\hat{f}_t = \frac{\lambda_0(t) + \sum_{n=1}^{\infty} I_t^{(n)}(\lambda_n)}{1 + \sum_{n=1}^{\infty} I_t^{(n)}(k_n)} \tag{4.7}$$

and the infinite series in (4.7) converge in the $L^1(P)$ norm.

Remarks 1. We call (4.6) the partial filter expansion and (4.7) the full expansion.

2. The expansions are essentially series evaluations of the Bayes formula (4.2). They work by "separating" the $x(\cdot)$ and $y(\cdot)$ dependence in (4.2) by expanding L_t; $y(\cdot)$ appears only when integrated in multiple integrals and $x(\cdot)$ is integrated out in the kernels. The kernels, therefore, require only knowledge of the apriori distribution of $x(\cdot)$ and can be computed offline prior to filtering.

Theorem 4.1 has the obvious generalization to vector valued processes.

The proof of theorem 4.1 requires that we handle integrals of the form $L_t^{(r)}$ or

$$\int_0^t \cdots \int_0^{s_{r-1}} h(s_1)\cdots h(s_r)dy(s_r)\cdots dy(s_1) \tag{4.8}$$

whose kernels are random, not deterministic as in chapter 3. These are easily defined by iteration. For this, it is convenient to work with the measure P_0, with respect to which $(y(t),G_t)$ is a Brownian process, so that we may apply the standard theory of stocnastic integration, (see, e.g., Liptser and Shiryayev [28]). Thus, if $b(t)$ is a measurable, G_t-adapted process satisfying $P_0[\int_0^T b^2(s)ds < \infty] = 1$, we have an integral $\int_0^t b(s)dy(s)$ with a version such that $P_0[\sup_{0\le t\le T}\int_0^t b(s)dy(s) < \infty] = 1$ (Liptser, Shiryayev [28]). The hypothesis of (4.1), $E\int_0^T h^2(s)ds < \infty$ is thus

enough to guarantee that (4.8) is well defined for all orders r. Indeed, $\psi(t) = \int_0^t h(s)dy(s)$ is certainly well defined, and, moreover,

$$P_0[\int_0^T h^2(s)\psi^2(s)ds \leq (\sup_{[0,T]} |\psi(s)|)^2 \int_0^T h^2(s)ds < \infty] = 1.$$

Hence

$$\int_0^t h(s_1)\psi(s_1)dy(s_1) = \int_0^t \int_0^{s_1} h(s_1)h(s_2)dy(s_2)dy(s_1)$$

is well defined, and we can continue in this manner to all orders. Similarly, by choosing a continuous version of L_t, we can show

$$L_t^r = \int_0^t \cdots \int_0^{s_r} h(s_1)\cdots h(s_r)L_{s_{r+1}} dy(s_r)\cdots dy(s_1)$$

is well-defined.

We shall also encounter expressions of the form

$$E_0(\int_0^t \phi(s)dy(s)|F_t^y),$$

and for these, the following "stochastic Fubini" theorem is useful.

Lemma 4.1 Let $(b(t),F_t)$ be a standard Wiener process and let $F_t^b = \sigma(b(s)|s\leq t)$ (completed by null sets). If $\phi(s)$ is an F_t-adapted process such that $E[\int_0^T \phi^2(s)ds] <$

$$E\left[\int_0^t \phi(s)db(s) \mid F_t^b\right] = \int_0^t E[\phi(s) \mid F_s^b]db(s) \quad .$$

Proof Liptser and Shiryayev [28].

Proof of Theorem 4.1 Parts (i) and (ii) are both consequences of a multiple integral expansion for L_t. Indeed, (4.3) implies that

$$L_t = 1 + \int_0^t h(s)L_s dy(s) \quad . \tag{4.9}$$

Iterating (4.9),

$$L_t = 1 + \int_0^t h(s)dy(s) + \int_0^t \int_0^{s_1} h(s_1)h(s_2)L_{s_2} dy(s_2)dy(s_1)$$

$$\vdots$$

$$L_t = 1 + \int_0^t h(s)dy(s) + \cdots + \int_0^t \cdots \int_0^{s_{r-1}} h(s_1)\cdots h(s_r)dy(s_r)\cdots dy(s_1)$$

$$+ L_t^{(r)} \quad . \tag{4.10}$$

Now substitute this last expression into

$$\hat{f}_t = \frac{E_0(f(t)L_t \mid F_t^y)}{E_0(L_t \mid F_t^y)} \quad .$$

The denominator, for example, becomes

$$E_0\{L_t|F_t^y\} = 1 + \sum_{n=1}^{r} E_0\{\int_0^t \cdots \int_0^{s_{n-1}} h(s_1)\cdots h(s_n)dy(s_n)\cdots dy(s_1)|F_t^y\}$$

$$+ E_0\{L_t^{(r)}|F_t^y\} . \qquad (4.11)$$

The hypothesis $E[\int_0^T h^2(s)ds]^r < \infty$ of part (i) allows lemma 4.1 to be applied to the terms of (4.11), with the result,

$$E_0[L_t|F_t^y] = 1 + \sum_{n=1}^{r} \int_0^t \cdots \int_0^{s_{n-1}} E_0\{h(s_1)\cdots h(s_n)\}dy(s_n)\cdots dy(s_1) +$$

$$E_0\{L_t^{(r)}|F_t^y\} = 1 + \sum_{n=1}^{r} I_t^n(k_n) + E_0\{L_t^{(r)}|F_t^y\}.$$

A similar calculation applies to $E_0[f(t)L_t|F_t^y]$, thereby completing the derivation of the partial expansion.

Formally, the proof of the full expansion follows by setting $r = \infty$ in (4.10). To prove it rigorously, we first show that $E \exp[\int_0^T h^2(s)ds] < \infty$ implies

$$L_t = \text{m.s.}(p_0) \lim_{N\to\infty}[1 + \sum_{n=1}^{N} \int_0^t \cdots \int_0^{s_{n-1}} h(s_1)\cdots h(s_n)dy(s_n)\cdots dy(s_1)]. \qquad (4.12)$$

Denote the finite series on the right hand side of (4.12) by L_t^N. Then

$$E_0(L_t - A_t^N)^2 = E_0\left[\int_0^t \cdots \int_0^{s_N} h(s_1)\cdots h(s_{N+1}) L_{s_{N+1}}\, dy(s_{N+1})\cdots dy(s_1)\right]^2.$$

By employing the standard computational rules of (2.1) for stochastic integrals, this last expression equals

$$\int_0^t \cdots \int_0^{s_N} E_0[h^2(s_1)\cdots h^2(s_N) L^2_{s_{N+1}}]\, ds_{N+1}\cdots ds$$

provided that it is finite. However,

$$\begin{aligned} &E_0[h^2(s_1)\cdots h^2(s_N) L^2_{s_{N+1}}] \\ &= E_0\{h^2(s_1)\cdots h^2(s_N)\, \exp[-\int_0^{s_{N+1}} h^2(s)ds] \qquad (4.13) \\ &\quad E_0[\exp[2\int_0^{s_{N+1}} h(s)dy(s)]|F^x_{s_1}]\}. \end{aligned}$$

With respect to P_0, $x(\cdot)$ and $y(\cdot)$ are independent and $y(\cdot)$ is Brownian. Hence, given $\{x(s),\ s \le s_1\}$, $\int_0^{s_{N+1}} h(s)dy(s)$ is a Gaussian random variable with mean 0 and variance $\int_0^{s_{N+1}} h^2(s)ds$. Thus

$$E_0[\exp 2\int_0^{s_{N+1}} h(s)dy(s)|F^x_{s_1}] = \exp 2\int_0^{s_{N+1}} h^2(s)ds\ . \qquad (4.14)$$

Therefore, applying (4.14) to (4.13)

$$(4.13) = E_0\{h^2(s_1)\cdots h^2(s_{N+1})\exp[\int_0^{s_{N+1}} h^2(s)ds]\}$$

$$= E_0\{h^2(s_1)\cdots h^2(s_{N+1})\sum_{j=N+1}^{\infty}\int_0^{s_{N+1}}\int_0^{\sigma_1}\cdots\int_0^{\sigma_{j-1}} h^2(\sigma_1)\cdots h^2(\sigma_j)d\sigma_j\cdots d\sigma_1\}$$

and hence

$$\int_0^t\cdots\int_0^{s_{N+1}} E_0[h^2(s_1)\cdots h^2(s_{N+1})L^2_{s_{N+1}}]ds_{N+1}\cdots s_1$$

$$= \sum_{j=N+1}^{\infty}\int_0^t\cdots\int_0^{s_{j-1}} E_0[h^2(s_1)\cdots h^2(s_j)]ds_j\cdots ds_1$$

$$= \sum_{j=N+1}^{\infty}\frac{1}{j!}E_0\{\int_0^T\cdots\int_0^T h^2(s_1)\cdots h^2(s_j)ds_j\cdots ds_1\} . \quad (4.16)$$

Since $E\exp[\int_0^T h^2(s)ds] < \infty$, (4.15) tends to 0 as $N \to \infty$, and thus $L_t = m.s\ (P_0)\lim_{N\to\infty} A_t^N$ for all $t \le T$. Lemma 4.1 can now be invoked for every order n, so that

$$E_0\{L_t|F_t^y\} = E_0\{ms.\lim_{N\to\infty} A_t^N|F_t^y\}$$

$$= m.s.\lim_{N\to\infty} E_0\{A_t^N|F_t^y\}$$

$$= m.s.(P_0)\lim_{N\to\infty}[1 + \sum_{n=1}^{N} I_n^n(<\ldots>)] .$$

A similar proof expands $E_0(f(t)L_t|F_t^y)$ in the series

$$\alpha_0(t) + \sum_{n=1}^{\infty} I_t^n(\alpha_n) .$$

Finally, to derive the $L^1(P)$ convergence, note that

$$E_0\left[\frac{dP}{dP_0}\right]^2 = E_0 L_T^2 = E\left[\exp \int_0^T h^2(s)ds\right] < \infty .$$

Thus

$$E\left|E_0[L_t|F_t^y] - \left(1 + \sum_1^N I_t^n(k_n)\right)\right|$$

$$\leq E_0^{1/2}\left(\frac{dP}{dP_0}\right)^2 E_0^{1/2}\left[E_0[L_t|F_t^y] - \left(1 + \sum_1^N I_t^n(k_n)\right)\right]^2 .$$

Thus because of (4.12), $E_0[L_t|F_t^y] = (L^1(P))\lim_{N\to\infty}\left[1 + \sum_{n=1}^{N} I_t^n(k_n)\right]$ as claimed. This completes the proof of theorem 4.1.

Let $P(\Delta,t|F_t^y) = E[1_\Delta(x(t))|F_t^y]$ denote the conditional distribution of $x(t)$ given the observation up to time t.

<u>Corollary</u> If $E\left[\exp \int_0^T h^2(s)ds\right] < \infty$

$$P(\Delta,t|F_t^y) = \frac{E1_\Delta(x(t)) + \sum_{n=1}^{\infty} I_t^n(E1_\Delta(x(t))h(s_1)\cdots h(s_n))}{1 + \sum_{n=1}^{\infty} I_t^n(Eh(s_1)\cdots h(s_n))} .$$

A related formula is also of interest. If $x(t)$ has a density $q(x,t)$, $x(t)$ has a conditional density given by

$$p(x,t)|F_t^y) = \frac{E_0(L(t)|F_t^y, x(t)=x]q(x,t)}{E_0[L(t)|F_t^y]} \quad .$$

Using the same techniques as above, we can easily derive

$$E_0[L(t)|F_t^y, x(t)=x]q(x,t) = [1 + \sum_{n=0}^{\infty} I_t^n(E[h(s_1)\cdots h(s_n)|x(t)=x])] \times q(x,t) \qquad (4.16)$$

for the numerator of $p(x,t|F_t^y)$. (4.16) is often called the unnormalized conditional density.

Theorem 4.1 immediately suggests a scheme for approximating filters, namely, truncation of the numerator and denominator series after a finite number of terms. The kernels of these terms are evaluated off-line, and, if necessary, approximated by separable versions. Construction of the multiple integrals as outputs of stochastic differential equations in the manner of theorem 2.3, then provides a finite-dimensional recursive realization of the approximate filter. Error analysis of this method is difficult, even in the case $E[\exp \int_0^T h^2(s)ds] < \infty$; because truncation occurs in both numerator and denominator of a ratio and because the error terms are hard to bound. One might also approximate the unnormalized conditional $p(x,t|F_t^y)$ by a finite series, but a major drawback to this scheme is that one cannot guarantee that $p(x,t|F_t^y) \geq 0$ for all x.

An alternative is to discuss cumulant expansions

$$p(x,t|\mathcal{F}_t^y) = \exp \sum_{n=0}^{\infty} I_t(a_n) \quad .$$

Eterno [11] studies ideas like this in his thesis.

4.2 Best r^{th} Order Filters

The most common and extensively studied suboptimal estimator is, of course, the best linear filter. This is an estimator

$$\tilde{f}_t = a_0 + \int_0^t a_1(t,s)dy(s),$$

linear in $y(\cdot)$, and satisfying

$$E[f(t)-\tilde{f}_t]^2 \leq E[f(t)-(b_0+\int_0^t b_1(t,s)dy(s))]^2$$

for all other choices of b_0 and $b_1(t,s)$. The philosophy of applying multiple integrals naturally suggests that one seek better-than-linear estimates by adding higher order multiple integrals terms.

Definition 4.1

(i) $Y_r \equiv \{a(t)=a_0(t) + \sum_{n=1}^{r} I_t^n(a_n(t)), t\leq T, a_n(t,s_1,\cdots,s_n)\in\hat{L}^2([0,T]^n), t\leq T$

and $0 \leq n \leq r\}$

(ii) $a(t) \in Y_r$ is called the best r^{th} order filter of $f(t)$, given $\{y(s) | s \leq t\}$, if $E[f(t)-a(t)]^2 \leq E[f(t)-b(t)]^2$ for $t \leq T$, for all $b(t) \in Y_r$. The kernels $a_j(t)$ of $a(t)$ are then called the optimal kernels.

Existence of a best r^{th} order filter is not immediately guaranteed. $a(t)$ amounts to the projection of $f(t)$ upon $Y_r(t) = \{a(t) | a(\cdot) \in Y_r\}$, and, for this to always exist, $Y_r(t)$ must be closed under mean-square (P) limits. An easy sufficient condition is

Lemma 4.2 If $E(\int_0^T h^2(s)ds)^r < \infty$, then $Y_r(t)$ is mean-square (P) closed for $t \leq T$.

Proof Apply lemma 2.3 to observe that, under the hypothesis,

$$E[I_T^n(k)]^2 \leq M_n \|k\|^2$$

for $n \leq r$, $k \in \hat{L}^2([0,T]^n)$.

To find the best r^{th} order filter, one must compute the optimal kernels. Accordingly, in theorem 4.2 we show how to use the multiplication formula and the filter expansions to derive integral equations for the $a_n(t)$, $0 \leq n \leq r$. This requires two preliminary lemmas, one to restate the problem, the other, to verify a technical identity. We assume throughout the notations established in §1.

Lemma 4.3 Let $z,v \in L^2(\Omega,F_t^y,P)$. Then $E(z-f(t))^2 \leq E(v-f(t))^2$ if and only if $E(z-\hat{f}(t))^2 \leq E(v-\hat{f}(t))^2$.

Proof $\hat{f}(t)$ equals the projection of $f(t)$ onto $L^2(\Omega,F_t^y,P)$. Since $z-\hat{f}(t) \in L^2(\Omega,F_t^y,P)$, the projection theorem implies

$$E(f(t)-\hat{f}(t))\,(z-\hat{f}(t)) = 0 \quad .$$

Thus

$$E(z-f(t))^2 = E(z-\hat{f}(t))^2 + E(f(t)-\hat{f}(t))^2 - 2E(f(t)-\hat{f}(t))(z-\hat{f}(t))$$

$$= E(z-\hat{f}(t))^2 + E(f(t)-\hat{f}(t))^2$$

Similarly,

$$E(v-f(t))^2 = E(v-\hat{f}(t))^2 + E(f(t)-\hat{f}(t))^2 \quad .$$

Lemma 4.4 Let $c(t) = c_0(t) + \sum_{n=1}^{r} I_t^n(c_n(t)) \in Y_r(t)$ and assume that $E(\int_0^T h^2(s)ds)^r < \infty$, $E\, f^2(t)(\int_0^T h^2(s)ds)^r < \infty$. Then

$$E_0\{c(t)E_0[L_t^{(r)}|F_t^y]\} = 0 \tag{4.17}$$

$$E_0\{c(t)\,E_0[f(t)L_t^{(r)}|F_t^y]\} = 0 \tag{4.18}$$

Proof From (4.11)

$$E_0\{L_t^{(r)}|F_t^y\} = E_0[L_t|F_t^y] - 1 - \sum_{n=1}^{r} I_t^n(k_n),$$

and, therefore,

$$E_0\{c(t)E_0\{L_t^{(r)}|F_t^y\}\} = E_0c(t)E_0[L_t|F_t^y] - E_0c(t)[1 + \sum_{n=1}^{r} I_t^n(k_n)] \tag{4.19a}$$

Since $y(\cdot)$ is Brownian under P_0

$$E_0\{c(t)[1 + \sum_{n=1}^{r} I_t^n(k_n)]\} = c_0(t) + \sum_{n=1}^{r} \int_0^t \cdots \int_0^{s_{n-1}} c_n(t,s_1,\cdots,s_n) E[h(s_1)\cdots h(s_n)]ds_n\cdots ds_1 \tag{4.19b}$$

However,

$$E_0\{c(t)E_0[L_t|F_t^y]\} = E_0c(t)\frac{dP}{dP_0} = Ec(t)$$

$$= c_0(t) + \sum_{n=1}^{r} \int_0^t \cdots \int_0^{s_{n-1}} c_n(t,s_1,\cdots,s_n) E[h(s_1)\cdots h(s_n)]ds_n\cdots ds_1 \tag{4.20}$$

by lemma 2.3. Applying (4.19b) and (4.20) in (4.19a) yields

$$E_0\{c(t)E_0[L_t|F_t^y]\} = 0 \quad .$$

(4.18) is established in analogous fashion using a version of lemma 2.3 for expressions $f(t)\, I_t^n(c_n)$, $n \le r$, under the condition $E\{f^2(t)(\int_0^t h^2(s)ds)^r\} < \infty$.

Suppose now that $E(\int_0^T h^2(s)ds)^{2r}$ and $a(\cdot) \in Y_r$. By expanding $E_0[L_t|F_t^y]$ as in theorem 4.1, we obtain

$$a(t)\, E_0[L_t|F_t^y] = [a_0(t) + \sum_{n=1}^{r} I_t^n(a_n(t))]$$

$$\times [1 + \sum_{j=1}^{2r} I_t^j(k_j) + E_0\{L_t^{(2r)}|F_t^y\}] . \qquad (4.21)$$

Using the multiplication formula, we can then calculate kernels $g_j(t,s_1,\cdots,s_j)$ such that

$$(4.21) = g_0(t) + \sum_{j=1}^{3r} I_t^j(g_j(t,s_1,\cdots,s_j)) + a(t)\, E_0\{L_t^{(2r)}|F_t^y\}. \qquad (4.21a)$$

Indeed, for $0 \le j \le 3r$, direct calculation with (2.14) gives

$$g_j = \sum_{(m,n,i)\in C_j} \binom{m+n-2i}{m-i} [a_m(t)]_i \odot (t)\, [k_n]_i \qquad (4.22)$$

$$C_j = \{m,n,i \mid m+n-2i=j, i\le\min(m,n), m\le r, n\le 2r\}$$

Theorem 4.2 Assume $E[\int_0^T h^2(s)ds]^{2r} < \infty$ and $E\{a^2(t)[\int_0^T h^2(s)ds]^{2r}\} < \infty$.

$a(t)$ is the best r^{th} order estimate if and only if

$$g_j(t,s_1,\cdots,s_j) = E\{f(t)h(s_1)\cdots h(s_j)\}\ (= \alpha_j(t,s_1,\cdots,s_j)) \qquad (4.23)$$

for $1 \leq j \leq r$.

Remark The equations (4.23) comprise $r+1$ integral equations for the $r+1$ kernels $a_j(t)$, $0 \leq j \leq r$. This can be seen from (4.23) and the definition of $\odot$ and will be illustrated explicitly in the examples to be discussed.

Proof Because of lemma 4.3 it suffices to show (4.23) holds if and only if

$$E[a(t)-\hat{f}(t)]^2 \leq E[c(t)-\hat{f}(t)]^2$$

for all $c(t) \in Y_r(t)$. Since

$$E[c(t)-\hat{f}(t)]^2 = E[c(t)-a(t)]^2 + E[a(t)-\hat{f}(t)]^2 + 2E[c(t)-a(t)]$$

$$\times[a(t)-\hat{f}(t)]$$

this will occur if and only if

$$E[c(t)-a(t)][a(t)-\hat{f}(t)] = 0 \quad \forall\, c(t) \in Y_r(t) \qquad (4.24)$$

Thus, we will demonstrate (4.24). Begin by noting that

$$E[\frac{dP_0}{dP}|F_t^y] = (E_0[\frac{dP}{dP_0}|F_t^y])^{-1} = (E_0[L_t|F_t^y])^{-1} .$$

Then

$$E[c(t)-a(t)][a(t)-\hat{f}(t)] = E\{\frac{(c(t)-a(t))[a(t)E_0[L_t|F_t^y]-E_0[f(t)L_t|F_t^y]]}{E_0[L_t|F_t^y]}\}$$

$$= E\{E[\frac{dP}{dP_0}|F_t^y](c(t)-a(t))[a(t)E_0[L_t|F_t^y] - E_0[f(t)L_t|F_t^y]]\}$$

$$= E_0\{(c(t)-a(t))(a(t)E_0[L_t|F_t^y] - E_0[f(t)L_t|F_t^y])\} \quad . \quad (4.25)$$

Now use theorem 4.1 (i) and (4.21a) to evaluate the term

$$a(t)E_0[L_t|F_t^y] - E_0[f(t)L_t|F_t^y]$$

$$= g_0(t)-Ef(t) + \sum_{j=1}^{2r} I_t^j[g_j(t)-a_j(t)] -$$

$$\sum_{j=2r+1}^{3r} I_t^j[g_j(t)]$$

$$+ a(t)E_0[L_t^{(2r)}|F_t^y] - E_0[f(t)L_t^{(2r)}|F_t^y].$$

This implies, since $y(\cdot)$ is Brownian w.r.t. P_0, that different index integrals are orthogonal w.r.t. P_0, and that, hence,

$$(4.25) = [c_0(t)-a_0(t)][g_0(t)-\alpha_0(t)] +$$

$$+ \sum_{j=1}^{r} \int_0^t \cdots \int_0^{s_{j-1}} [c_j(t,s_1,\cdots,s_j) - a_j(t,s_1,\cdots,s_j)]$$

$$[g_j(t,s_1,\cdots,s_j) - \alpha_j(t,s_1,\cdots,s_j)]$$

$$ds_j \cdots ds_1$$

$$+ E_0\{(c(t)-a(t))a(t)E_0[L_t^{(2r)} | F_t^y]\}$$

$$- E_0\{(c(t)-a(t))E_0[f(t)L_t^{(2r)} | F_t^y]\} \quad , \qquad (4.26)$$

The last two terms of (4.26) are zero by lemma 4.3. Thus, it is clear that (4.26) is zero iff

$$g_j = \alpha_j \qquad 0 \le j \le r \quad .$$

This completes the proof.

The technique of theorem 4.2 extends to other problems as well. Suppose, for instance, that a filter

$$a'(t) = a_0'(t) + \sum_{j=1}^{q} I_t^q(a_j'(t))$$

of order q is available; $a'(t)$ need not be the best q^{th} order filter. Let $r > q$, and, rather than ask for the best r^{th} order filter, let us seek the 'best r^{th} order correction' to $a'(t)$, i.e., the mean-square minimizing $a(t)$ of the form

$$a(t) = a_0^t(t) + \sum_{j=1}^{q} I_t^j(a_j^t(t)) + \sum_{j=q+1}^{r} I_t^j(a_j(t))$$

where $a_j(t)$, $j = q+1,\cdots,r$ are free to be chosen. Define the kernels $g_j(t)$ as before:

$$a(t)E_0\{L_t|F_t^y\} = g_0(t) + \sum_{j=1}^{3r} I_t^j(g_j(t)) + a(t)E_0\{L_t^{(2r)}|F_t^y\} .$$

Theorem 4.3 Let the hypotheses of theorem 4.2 hold. Then $a(t)$ is the best r^{th} order correction to $a'(t)$ if and only if

$$g_j(t,s_1,\cdots,s_j) = E\{f(t)h(s_1)\cdots h(s_j)\},\ q+1 \le j \le r. \quad (4.27)$$

Proof As before, it suffices to show that (4.27) holds iff

$$E[c(t)-a(t)][a(t)-\hat{f}(t)] = 0$$

for all $c(t) = a'(t) + \sum_{j=q+1}^{r} I_t^j(c_j(t))$. By the same calculations as in theorem 4.2

$$E[c(t)-a(t)][a(t)-\hat{f}(t)]$$

$$= E_0\{[c(t)-a(t)][a(t)E_0\{L_t|F_t^y\} - E_0\{f(t)L_t|F_t^y\}]\}$$

$$= E_0\{\sum_{j=q+1}^{r} I_t^j(c_j-a_j)[g_0(t)-a_0(t) + \sum_{j=1}^{2r} I_t^j[g_j(t)-a_j(t)]$$

$$+ \sum_{j=2^{q-1}}^{3r} I_t^j(g_j(t)) + a(t)E_0[L_t^{2r}|F_t^y] - E_0\{f(t)L_t^{(2r)}|F_t^y\}\}$$

$$= \sum_{j=q+1}^{r} \int_0^t \cdots \int_0^{s_{j-1}} [c_j-a_j](t,s_1,\cdots,s_j)[g_j-a_j](t,s_1,\cdots,s_j)\,ds_j\cdots ds_1$$

This equals zero iff $g_j = \lambda_j$ for $q - 1 \leq j \leq r$.

Remark Clearly, an analogous result holds for the case in which an arbitrary subset of $\{a_j\}_{j=0}^{r}$ is given and the remainder are chosen as to optimize the mean-square filter error. Thus, if a_j, $j \in \{j_1,\cdots,j_q\} \subseteq \{0,1,\cdots,r\}$ are given, then the $\{a_j(t)\}$, $j \notin \{j_1,\cdots,j_q\}$ are optimally chosen iff $g_j = \lambda_j$ for every $j \in \{0,1,\cdots,r\} - \{j_1,\cdots,j_q\}$.

As a first example of theorem 4.2, let us compute the kernel equations for the best linear estimate $\tilde{f}(t) = a_0(t) + \int_0^t a_1(t,s)dy(s)$. From (4.22),

$$g_0(t) = a_0(t) + \int_0^t a_1(t,\sigma)E[h(\sigma)]d\sigma$$

$$g_1(t,s) = a_1(t,s) + \int_0^t a_1(t,\sigma)E[h(s)h(\sigma)]d\sigma + a_0(t)Eh(s).$$

The kernel equations are then

$$a_0(t) + \int_0^t a_1(t,\sigma)Eh(\sigma)d\sigma = Ef(t)$$

$a_0(t)Eh(s) + a_1(t,s) + \int_0^t a_1(t,\sigma)E[h(s)h(\sigma)]d\sigma = Ef(t)h(s)$, or, eliminating $a_0(t)$ from the second equation,

$$a_0(t) + \int_0^t a_1(t,\sigma)E[h(\sigma)]d\sigma = Ef(t)$$

$$a_1(t,s) + \int_0^t a_1(t,\sigma)\mathrm{cov}[h(s),h(\sigma)]d\sigma = \mathrm{cov}[f(t),h(s)] \qquad (4.28)$$

(4.28) is, of course, the well-known Wiener-Hopf type equation for otpimal linear filtering. Before examining higher order examples, we will discuss the Kalman filter.

4.3 The Kalman Filter

Consider again the signal-observation system (3.1) in which the state $x(t)$ is a Gauss-Markov process solving a linear stochastic differential equation driven by Brownian motion, and $h(t,x) = H(t)x(t)$. The Kalman-Bucy theorem, summarized in lemma 3.1, shows that the state estimator $\hat{x}(t) = E(x(t)|\mathcal{F}_t^y)$ satisfies the equation

$$d\hat{x}(t) = F(t)\hat{x}(t)dt + P(t)H^T(t)[dy(t) - H(t)\hat{x}(t)dt]$$
$$\hat{x}(0) = x_0 \qquad (4.29)$$

where $P(t)$ is a deterministic function. (see chapter 3 for definitions.) Let $\Phi(t,s)$ be the state transition matrix of $F(t) - P(t)H^T(t)H(t)$. Then, the solution of (4.29) is

$$\hat{x}(t) = \Phi(t,0)x_0 + \int_0^t \Phi(t,s)P(s)H^T(s)dy(s),$$

i.e., the optimal estimate is linear in $y(\cdot)$.

It is of interest to connect this result to the expansion formulae of theorem 4.1 In the case of a scalar signal, satisfying (3.1) and h(t,x) = x,(4.7) yields, at least formally,

$$\hat{x}(t) = \frac{Ex(t) + \sum_{j=1}^{\infty} I_t^j(E[x(t)x(s_1)\cdots x(s_j)])}{1 + \sum_{j=1}^{\infty} I_t^j (E(x(s_1)\cdots x(s_j))} , \qquad (4.30)$$

and both numerator and denominator are truly infinite sums. This general representation obscures the linear structure of $\hat{x}(t)$. The techniques for applying the expansion formulae, should at least include methods for deriving the linearity of $\hat{x}(t)$ from (4.30). In fact, theorem 4.2 can be parlayed into a proof of the Kalman-Bucy filter and we present this here after a few comments.

One common proof of the Kalman-Bucy theorem invokes the stochastic differential equations for the conditional moments. When x(t) = b(t) and h(t,x) = x, where b(t) is a scalar Brownian motion, these are

$$d\hat{x^n}(t) = \frac{n(n-1)}{2} \widehat{x(t)^{n-2}}\, dt - [\widehat{x(t)^{n+1}} - \hat{x}(t)\hat{x^n}(t)]$$

$$[dy(t) - \hat{x}(t)dt] \quad n \geq 1 \qquad (4.31)$$

and a similar infinite set of coupled equations holds in the general case (3.1) (Fujisaki, Kallianpur, Kunita [15]). These equations do

AD-A096 973 MASSACHUSETTS INST OF TECH CAMBRIDGE LAB FOR INFORMA--ETC F/G 12/1
TOPICS IN NONLINEAR FILTERING THEORY.(U)
DEC 80 D L OCONE AFOSR-77-3281
UNCLASSIFIED LIDS-TH-1058 AFOSR-TR-81-0232 NL

2 OF 2
AD A 096973

END
DATE FILMED
4-81
DTIC

not yield themselves to a direct solution. Rather, they require additional information, namely, that $(x(t),y(t))$ is jointly Gaussian and that, hence, by limiting arguments, the conditional distribution of $x(t)$ given $\{y(s)|0 \leq s \leq t\}$ is normal. One can then conclude that the conditional variance $E(x-\hat{x})^2|\mathcal{F}_t^y]$ is a deterministic and so effectively truncate the system (4.31) at $n = 2$. (4.29) follows easily [see Kallianpur-Striebel [23]]. By contrast, the derivation of $\hat{x}(t)$ from (4.30) does not involve knowing the form of the conditional distribution, an object, that, in the general filtering problem, is not often in hand.

Let us develop our proof of the Kalman filter for the simple case

$$\begin{aligned} dx(t) &= db(t) \quad x(0) = 0 \\ dy(t) &= x(t)dt + dw(t) \quad y(0) = 0 \end{aligned} \tag{4.32}$$

in which $b(\cdot)$ is a Brownian motion. We do this in the interest of computational simplicity; the proof carries over to the general case.

<u>Theorem 4.4</u> $\hat{x}(t) = \int_0^t a(t,s)dy(s)$ where $a(t,s)$ satisfies the Wiener-Hopf equation $a(t,s) + \int_0^t a(t,\sigma)(s\wedge\sigma)d\sigma = s,\ t \geq s,$ (4.33) $(s\wedge\sigma = \min(s,\sigma))$.

<u>Proof</u> Since $y(\cdot)$ is Gaussian, the set of polynomials in $y(\cdot)$ is dense in $L^2(\Omega,\mathcal{F}_t^y,P)$, a fact presented in the discussion of homo-

geneous chaos theory in chapter 1. Therefore, it suffices to show that $\int_0^t a(t,s)dy(s)$ is the best r^{th} order estimate for every r, $1 \le r < \infty$. It is true for every r, $T < \infty$ and $t \le T$ that $E[\int_0^t b^2(s)ds]^r < \infty$ $E\, b^2(t)[\int_0^t b^2(s)ds]^r < \infty$. Theorem 4.2 thus applies. That is, $\int_0^t a(t,s)dy(s)$ is the best r^{th} order estimate if and only if

$$g_j(t,s_1,\cdots,s_j) = E\{b(t)b(s_1)\cdots b(s_j)\},\ j=0,1,\cdots,r\ .$$

From (4.22)

$$g_0(t) = 0$$

$$g_j(t,s_1,\cdots,s_j) = j(a(t,\cdot)\odot(t)k_{j-1})(s_1,\cdots,s_j)$$

$$+ (a(t,\cdot)\odot(t)(k_{j+1})_1)(s_1,\cdots,s_j)\quad j > 0.$$

Now

$$j(a(t,\cdot)\odot(t)k_{j-1})(s_1,\cdots,s_j) = \frac{j}{j!}\sum_{\pi\in S_j} a(t,s_{\pi(1)})$$

$$Eb(s_{\pi(2)})\cdots b(s_{\pi(j)})$$

$$= \sum_{i=1}^{j} a(t,s_i)E\{b(s_1)\cdots b(s_{i-1})b(s_{i+1})\cdots b(s_j)\}\ ,$$

and

$$(a(t,\cdot)_1\odot(t)(k_{j+1})_1(s_1,\cdots,s_j)$$

$$= \int_0^t a(t,\sigma)E\{b(\sigma)b(s_1)\cdots b(s_j)\}d\sigma\ .$$

The kernel equations become

$$0 = Eb(t) \tag{4.34}$$

$$\hat{a}(t,s) + \int_0^t a(t,\sigma)E\{b(\sigma)b(s)\}d\sigma = E[b(t)b(s)] \tag{4.35}$$

$$\sum_{i=1}^{j} a(t,s_i)E\{b(s_1)\cdots b(s_{i-1})b(s_{i+1})\cdots b(s_n)$$

$$+ \int_0^t a(t,\sigma)E\{b(\sigma)b(s_1)\cdots b(s_j)\}d\sigma$$

$$= E[b(t)b(s_1)\cdots b(s_j)] \tag{4.36}$$j

$$2 \leq j \leq r$$

(4.34) is true by definition, and (4.35) is just (4.33). We now claim that if $a(t,s)$ satisfies (4.35), $(4.36)_j$ is true for all $j \geq 2$. This will imply that the equations for $\int_0^t a(t,s)dy(s)$ being the best r^{th} order estimate are satisfied for every r, and will complete the proof. To do this, assume $a(t,s)$ solves 4.35, and observe the identity

$$E\{b(\sigma)b(s_1)\cdots b(s_j)\} = \sum_{i=1}^{j} (\sigma \wedge s_i)E[b(s_1)\cdots b(s_{i-1})b(s_{i+1})\cdots b(s_j)] \tag{4.37}$$

(4.37) results from a direct application of lemma 3.3.

Now substitute (4.37) into the left-hand side of $(4.36)_j$ and use the Wiener-Hopf equation for $a(t,s)$:

$$\sum_{i=1}^{j} a(t,s_i)E[b(s_1)\cdots b(s_{i-1})b(s_{i+1})\cdots b(s_j)]$$

$$+ \int_0^t a(t,\sigma)E[b(\sigma)b(s_1)\cdots b(s_j)]d\sigma$$

$$= \sum_{i=1}^{j} \{a(t,s_i) + \int_0^t a(t,\sigma)[\sigma \wedge s_i]d\sigma\}\ E[b(s_1)\cdots b(s_{i-1})b(s_{i+1})\cdots b(s_j)]$$

$$= \sum_{i=1}^{j} (s_i \wedge t)\ E[b(s_i)\cdots b(s_{i-1})b(s_{i+1})\cdots b(s_j)]$$

$$= E[b(t)b(s_1)\cdots b(s_j)] \quad .$$

The last equality employs lemma 3.3 again and validates (4.36)j for any j.

4.4 Quadratic Filters

In this section we treat best second order, or quadratic, filters as an example of the theory of 4.2. We first present the optimal kernel equations for this case and then show how they may be solved. To guarantee validity of the discussion, we assume throughout the hypotheses of theorem 4.2 for r = 2:

$$E(\int_0^T h^2(s)ds)^4 < \infty$$

$$Ef^2(t)(\int_0^T h^2(s)ds)^4 < \infty \qquad \forall t \leq T \ .$$

Deriving the integral equations is simply a matter of calculation. Let $a(t) = a_0(t) + \int_0^t a_1(t,s)dy(s) + \int_0^t\int_0^{s_1} a_2(t,s_1,s_2)dy(s_2)dy(s_1)$ be a quadratic estimate and let $g_j(t,s_1,\cdots,s_j)$, $0 \le j \le 2$, be the kernels associated with $a(t)$ in the manner of (4.21a). Thus, using (4.22)

$$g_0(t) = a_0(t) + [a_1(t,\cdot)]_1 \odot (t)[k_1]_1 + [a_2(t,\cdot)]_2 \odot(t)[k_2]_2$$

$$g_1(t,s) = a_1(t,s) + a_0(t)k_1(s) + [a_1(t,\cdot)]_1 \odot (t)[k_2]_1(s) \qquad (4.38)$$

$$+ [a_2(t,\cdot)]_1 \odot(t)[k_1]_1(s) + [a_2(t,\cdot)]_2 \odot(t)[k_3]_2(s)$$

$$g_2(t,s_1,s_2) = a_2(t,s_1,s_2) + a_0(t)k_2(s_1,s_2) + [a_1(t,\cdot)] \odot(t)[k_1](s_1,s_2)$$

$$+ [a_1(t)]_1 \odot (t)[k_3]_1\ (s_1,s_2) + 2[a_2(t)]_1 \odot(t)\ [k_2]_1(s_1,s_2)$$

$$+ [a_2(t)]_2 \odot (t)[k_4]_2(s_1,s_2).$$

By theorem 4.2, $a(t)$ is optimal quadratic if and only if

$$Ef(t) = g_0(t)$$

$$Ef(t)h(s) = g_1(t,s) \qquad (4.39)$$

$$Ef(t)h(s_1)h(s_2) = g_2(t,s_1,s_2) \quad .$$

Now evaluate g_j, $0 \le j \le 2$ in (4.39) using (4.38) and the definitions of $\odot(t)$ and k_j. The result is in its full blown

ugliness,

$$Ef(t) = a_0(t) + \int_0^t a_1(t,s)Eh(s)ds + \int_0^t\int_0^{s_1} a_2(t,s_1,s_2)Eh(s_1)h(s_2)ds_2ds_1 \tag{4.40a}$$

$$Ef(t)h(s) = a_1(t,s) + a_0(t)Eh(s) + \int_0^t a_1(t,\sigma)Eh(\sigma)h(s)d\sigma + \int_0^t a_2(t,s,\sigma)Eh(\sigma)d\sigma + \int_0^t\int_0^{\sigma_1} a_2(t,\sigma_1,\sigma_2)Eh(\sigma_1)h(\sigma_2)h(s)ds \tag{4.40b}$$

$$\begin{aligned} Ef(t)h(s_1)h(s_2) = {} & a_2(t,s_1,s_2) + a_0(t)Eh(s_1)h(s_2) + a_1(t,s_1)Eh(s_2) \\ & + a_1(t,s_2)Eh(s_1) + \int_0^t a_1(t,\sigma)Eh(\sigma)h(s_1)h(s_2)d\sigma \\ & + \int_0^t [a_2(t,s_1,\sigma)Eh(\sigma)h(s_2) + a_2(t,s_2,\sigma)Eh(\sigma)h(s_1)]d\sigma \\ & + \int_0^t\int_0^{\sigma_1} a_2(t,\sigma_1,\sigma_2)E\{h(s_1)h(s_2)h(\sigma_1)h(\sigma_2)\}\, d\sigma_2d\sigma_1 \end{aligned} \tag{4.40c}$$

These equations deserve some elementary remarks before we set about solving them. First, the optimal kernels are all interrelated in the general case. We cannot solve for a_0 and a_1 independently of knowing a_2. Likewise, if $a_0 = c_0$, $a_1 = c_1$ are the kernels of the best linear estimate, they will not, in general be the lower order kernels of the best quadratic estimate. Secondly, the equations (4.40) can be used for other suboptimal designs in the spirit of theorem 4.3. Thus, if a_0 and a_1 are given, and we

seek the best quadratic correction to $a_0(t) + \int_0^t a_1(t,s)dy(s)$, this will be found by solving (4.40c) for a_2 in terms of a_1 and a_0. The methods developed for solving the full set of equations (4.40) will also apply to this problem.

To solve (4.40), we first eliminate a_0 and a_1 to derive an integral equation solely for a_2. a_0 is easy to handle. Merely solve (4.40a) for $a_0(t)$ in terms of a_1, a_2, and the known functions $Ef(t)$, $Ef(t)h(s)$, etc. and substitute this expression in (4.40b) and (4.40c). To further simplify, use (4.40b) in $a_1(t,s_1)Eh(s_2) + a_1(t,s_2)Eh(s_1)$ of (4.40c). We thus derive

$$a_0(t) = Ef(t) - \int_0^t a_1(t,\sigma)Eh(\sigma)d\sigma - \int_0^t \int_0^{\sigma_1} a_2(t,\sigma_1,\sigma_2)EH(\sigma_1)h(\sigma_2)d\sigma_2 d\sigma_1 \tag{4.41a}$$

$$a_1(t,s) = \mathrm{cov}[f(t),h(s)] - \int_0^t \mathrm{cov}[h(s),h(\sigma)]a_1(t,\sigma)d\sigma$$

$$- \int_0^t Eh(\sigma)a_2(t,s,\sigma)d\sigma$$

$$- \int_0^t \int_0^{\sigma_1} \mathrm{cov}[h(s),h(\sigma_1)h(\sigma_2)]a_2(t,\sigma_1,\sigma_2)d\sigma_2 d\sigma_1 \tag{4.41b}$$

$$a_2(t,s_1,s_2) = \mathrm{cov}[f(t),h(s_1),h(s_2)] - \int_0^t \mathrm{cov}[h(s_1),h(s_2),h(\sigma)]a_1(t,\sigma)d\sigma$$

$$- \int_0^t [\mathrm{cov}[h(s_1),h(\sigma)]a_2(t,s_2\sigma) + \mathrm{cov}[h(s_2),h(\sigma)]a_2(t,s_1\sigma)d\sigma$$

$$- \int_0^t \int_0^{\sigma_1} \mathrm{cov}[h(s_1),h(s_2),h(\sigma_1)h(\sigma_2)]a_2(t,\sigma_1,\sigma_2)d\sigma_2 d\sigma_1 \tag{4.41c}$$

(In these expressions, $\mathrm{cov}[x_1,\cdots x_r] = E(x_1-Ex_1)\cdots(x_r-Ex_r)$.)

We have yet to eliminate a_1 from (4.41b) and (4.41c), but this requires some more notation and a bit of theory. Define the operator $R_1(t)$: $L^2([0,t]) \to L^2([0,t])$ with kernel $r_1(\sigma,\sigma) = \mathrm{cov}[h(s),h(\sigma)]$ by

$$(R_1(t)g)(s) = \int_0^t \mathrm{cov}[h(s),h(\sigma)]g(\sigma)d\sigma \ .$$

$R_1(t)$ appears in (4.41b) and (4.41c). In particular, (4.41b) may be rewritten as

$$[I+R_1(t)](a_1(t,))(s) = \mathrm{cov}[f(t),h(s)] - \int_0^t Eh(\sigma)a_2(t,s,\sigma)d\sigma - \int_0^t \int_0^{\sigma_1} \mathrm{cov}[h(s),h(\sigma_1)h(\sigma_2)]a_2(t,\sigma_1,\sigma_2)d\sigma_2 d\sigma_1 \ , \tag{4.42}$$

and thus, solving for a_1 in terms of a_2 requires inverting $I+R_1(t)$. Fortunately, this can always be done in an explicit way.

Lemma 4.5

i) $h(t)$ has a best linear estimate $\hat{h}(t) = \alpha_0(t) + \int_0^t \alpha_1(t,s)dy(s)$, $t \le T$. Without loss of generality, we take $\alpha_1(t,s) = 0$ for $T \ge s \ge t \ge 0$.

ii) $I+R_1(t)$ is invertible, and $[I+R_1(t)]^{-1} = 1-Q(t)$ where $Q(t)$ is the integral operator with kernel

$$q(t,s_1,s_2) = \alpha_1(s_1,s_2) + \alpha_1(s_2,s_1) - \int_0^t \alpha_1(\sigma,s_1)\alpha_1(\sigma,s_2)d\sigma, 0 \le s_1,s_2 \le t$$

Proof $E[\int_0^T h^2(s)ds]^4 < \infty$ guarantees that $\hat{h}(t)$ exists (lemma 4.2) and, as in (4.28)

$$\alpha_1(s_1,s_2) + \int_0^{s_1} \alpha_1(s_1,\sigma)\, \mathrm{cov}[h(s_2),h(\sigma)]d\sigma$$

$$= \mathrm{cov}[h(s_1),h(s_2)] \quad 0 \le s_2 \le s_1 \le T$$

ii) follows from results of Kailath and Geesey (Geesey [16], Chapter 3). These imply that, under the hypothesis $\int_0^T \mathrm{cov}[h(s),h(s)]ds < \infty$, which is certainly implied by $E[\int_0^T h^2(s)ds]^4 < \infty$, $I+R_1(t)$ is invertible and its inverse has the given form.

We now apply this lemma to solve (4.42) for $a_1(t,s)$.

$$a_1(t,s) = \mathrm{cov}[f(t),h(s)] - \int_0^t q(t,s,\sigma)\, \mathrm{cov}[f(t),h(\sigma)]d\sigma$$

$$- \int_0^t \int_0^t r'(t,s,\sigma_1,\sigma_2)a_2(t,\sigma_1,\sigma_2)d\sigma_2 d\sigma_1 \qquad (4.43)$$

where

$$r'(t,s,\sigma_1,\sigma_2) = \frac{1}{2}\ \mathrm{cov}[h(s),h(\sigma_1)h(\sigma_2)]$$

$$+ \frac{1}{2}\ [q(t,s,\sigma_2)Eh(\sigma_1) + q(t,s,\sigma_1)Eh(\sigma_2)]$$

$$+ \frac{1}{2}\ \int_0^t q(t,s,\sigma)\, \mathrm{cov}[h(\sigma),h(\sigma_1)h(\sigma_2)]d\sigma .$$

In deriving r', we took advantage of the symmetry of a_2 to symmetrize r'. Finally, we substitute (4.43) into (4.41c) to get

$$a_2(t,s_1,s_2) = F(t,s_1,s_2)$$
$$- \int_0^t [r_1(s_1,\sigma)a_2(t,s_2,\sigma) + r_1(s_2,\sigma)a_2(t,s_1,\sigma)]d\sigma$$
$$- \int_0^t \int_0^t r_2(t,s_1,s_2,\sigma_1,\sigma_2)a_2(t,\sigma_1,\sigma_2)d\sigma_2 d\sigma_1 \qquad (4.44)$$

where

$$F(t,s_1,s_2) = \text{cov}[f(t),h(s_1),h(s_2)]$$
$$- \int_0^t \text{cov}[h(s_1),h(s_2),h(\sigma)](\text{cov}[f(t),h(\sigma)]$$
$$- \int_0^t q(t,\sigma_1\sigma_2)\,\text{cov}\,[f(t),h(\sigma_2)]d\sigma_2)d\sigma$$

$$r_2(t,s_1,s_2,\sigma_1,\sigma_2) = \frac{1}{2}\ [\text{cov}[h(s_1),h(s_2),h(\sigma_1),h(\sigma_2)]$$
$$- \text{cov}[h(s_1),h(s_2)]\,\text{cov}\,[h(\sigma_1),h(\sigma_2)]]$$
$$- \int_0^t \text{cov}[h(s_1),h(s_2),h(\tau)]\,r'(t,\tau,\sigma_1,\sigma_2)d\tau\ .$$

We have shown that if (a_0,a_1,a_2) solve (4.40) then a_2 solves (4.44). Conversely, by reversing the steps of this derivation, if a_2 solves (4.44) and a_0 and a_1 are defined via (4.43) and (4.41a), then a_0, a_1, a_2 solve (4.40).

(4.44) is simply a linear integral for a_2. However, the middle term of (4.44), involving a tensor contraction between a_2 and r_1, is non-standard, and the usual linear integral equation techniques do not directly apply. In what follows, we will show now

to eliminate the tensor contraction term to derive a Fredholm integral equation for a_2, thus reducing the kernel equations to a well-known problem for which methods of solution or approximate solution are readily available.

It is useful to do this in a general context.

Definition 4.2. Let $\beta(s_1,s_2)\in L^2([0,t]^2)$ and let $\gamma(s_1,s_2,\sigma_1,\sigma_2)\in L^2([0,t]^4)$ such that γ is symmetric in s_1,s_2. The operator Γ: $\hat{L}^2([0,t]^2) \to \hat{L}^2([0,t]^2)$

$$(\Gamma c)(s_1s_2) = [\beta]_1\odot(t)[c]_1(s_1,s_2) + \int_0^t\int_0^t \gamma(s_1,s_2,\sigma_1,\sigma_2)c(\sigma_1,\sigma_2)d\sigma_2d\sigma_1$$

is said to be of tensor contraction (T-C) type. The kernel β can also be used to define an integral operator on $L^2([0,t])$, which we will denote by B, and, in fact, we can write

$$[\beta]_1\odot(t)[c]_1(s_1,s_2) = \int_0^t [\beta(s_1,\sigma)c(s_2,\sigma) + \beta(s_2,\sigma)c(s_1,\sigma)]d\sigma$$

$$= (Bc(s_2,\cdot))(s_1) + (Bc(s_1,\cdot))(s_2) .$$

Remark: It is of interest to note that, while the second term of Γ is the usual compact, Hilbert-Schmidt operation, the tensor contraction term is not compact in general. Since compact operators have finite dimensional eigenspaces, we can prove this by supplying a β such that $\{c\in\hat{L}^2([0,t]^2) : c = \beta_1\odot(t)c_1\}$ is infinite dimensional. It is easily seen that

$$\beta(s_1,s_2) = \sum_{i=1}^{M} \phi_i(s_1)\phi_i(s_2) ,$$

where the ϕ_i , $1 \le i \le M < \infty$ are mutually orthogonal, will work. Then if $\{\psi_i\}_{i=1}^{M}$ are functions orthogonal to $\mathrm{Span}\{\phi_1,\cdots,\phi_M\}$, $c(s_1,s_2) = \sum_{i=1}^{M} [\phi_i(s_1)\psi_i(s_2) + \phi_i(s_2)\psi_i(s_1)]$ satisfies $c = \beta_1\cdot(t)c_1$. The space of such solutions is clearly infinite dimensional.

For a T-C operator Γ characterized by kernels β and γ, we want to solve the integral equation

$$c = F + \Gamma c \qquad , F \in \hat{L}^2([0,t]^2) \tag{4.45}$$

which generalizes (4.44).

Theorem 4.5. Suppose that I-B is invertible and

$$([I-B]^{-1}\phi)(s) = \phi(s) + \int_0^t \beta^1(s,\sigma)\phi(\sigma)d\sigma \qquad \beta^1 \in L^2([0,t]^2)$$

Then, if $c = \Gamma c$ has no non-zero solution, $I-\Gamma$ has a bounded inverse.

Proof Using $(I-B)^{-1}$ we will derive a Fredholm equation for c from (4.45). Thus, suppose $c = F + \Gamma c$. From the definition of Γ, we have

$$[(I-B)c(s_1,\cdot)](s_2) = c(s_1,s_2) - \int_0^t \beta(s_2,\sigma)c(s_1,\sigma)d\sigma$$

$$= F(s_1,s_2) + \int_0^t \beta(s_1,\sigma)c(s_2,\sigma)d\sigma$$

$$+ \int_0^t \int_0^t \gamma(s_1,s_2,\sigma_1,\sigma_2)c(\sigma_1\sigma_2)d\sigma_2 d\sigma_1$$

By inverting $I-B$ and rearranging terms, we derive

$$[(I-B)c(\cdot,s_2)](s_1) = F(s_1,s_2) + \int_0^t \beta^1(s_2,\sigma)F(s_1,\sigma)d\sigma$$

$$+ \int_0^t \int_0^t \gamma_1(s_1,s_2,\sigma_1,\sigma_2)c(\sigma_1,\sigma_2)d\sigma_2 d\sigma_1 \qquad (4.46)$$

where

$$\gamma_1(s_1,s_2,\sigma_1,\sigma_2) = \gamma(s_1,s_2,\sigma_1,\sigma_2) + \beta^1(s_2,\sigma_1)\beta(s_1,\sigma_2)$$

$$+ \int_0^t \beta^1(s_2,\eta)\ (s_1,\eta,\sigma_1,\sigma_2)d\eta$$

Now invert (4.46), to get

$$c(s_1,s_2) = F_1(s_1,s_2) + \int_0^t \int_0^t \gamma_3(s_1,s_2,\sigma_1,\sigma_2)c(c_1,\sigma_2)d\sigma_2 d\sigma_1 \qquad (4.47)$$

where

$$F_1(s_1,s_2) = F(s_1,s_2) + \int_0^t \beta^1(s_2,\sigma)F(s_1,\sigma)d\sigma$$

$$+ \int_0^t \beta^1(s_1,\sigma_1)[F(\sigma_1,s_2) + \int_0^t \beta^1(s_2,\sigma_2)F(\sigma_1,\sigma_2)d\sigma_2]d\sigma_1$$

and

$$\gamma_3(s_1,s_2,\sigma_1,\sigma_2) = \gamma_1(s_1,s_2,\sigma_1,\sigma_2) + \int_0^t \beta^1(s_1,\eta)\gamma_1(\eta,s_2,\sigma_1,\sigma_2)d\eta$$

(4.47) is the desired Fredholm equation for c ; if c solves $c = F + \Gamma c$ then c satisfies (4.47). Conversely, if c satisfies (4.47) then $c = F + \Gamma c$. Analogous reasoning shows that $c = \Gamma c$ iff

$$c(s_1,s_2) = (\Gamma_3 c)(s_1,s_2) = \int_0^t \int_0^t \gamma_3(s_1,s_2,\sigma_1,\sigma_2)d\sigma_2 d\sigma_1 \qquad (4.48)$$

Thus, if $c = \Gamma c$ has no non-zero solution, neither does $c = \Gamma_3 c$. But Γ_3 is a compact operator, and hence the Fredholm alternative theorem implies that $I-\Gamma_3$ is invertible. This clearly implies that $I-\Gamma$ is invertible also.

The statement of theorem 4.5 may seem odd because it does not focus on the central equation (4.47). This is done to emphasize that once the Fredholm equation is derived, we need conditions to guarantee it can be solved. Stipulating that 1 not be an eigenvalue of Γ provides just such a criterion. Further, the statement of theorem 4.5 may be extended to a sort of mutant Fredholm alternative for tensor contraction operators.

Corollary. If $I-\lambda B$ is invertible and λ is not an eigenvalue of Γ, then $I-\lambda\Gamma$ is invertible.

Proof. Completely analogous to that of theorem 4.5.

The equation (4.44) for $a_2(t,s_1,s_2)$ is of tensor contraction type; in fact, we may write it

$$a_2(t,s_1,s_2) = F(t,s_1,s_2) + [R(t)a_2(t,\cdot)](s_1,s_2) \qquad (4.49)$$

where $R(t)$ is the tensor contraction operator characterized by the kernels $\beta(\sigma_1,\sigma_2) = -r(\sigma_1,\sigma_2)$ and $\gamma(s_1,s_2,\sigma_1,\sigma_2) = -r_2(t,s_1,s_2,\sigma_1,\sigma_2)$.

Theorem 4.6. $I-R(t)$ is invertible and $a_2(t,\cdot,\cdot)$ is the unique solution of

$$a_2(t,s_1,s_2) = F_1(t,s_1,s_2) + \int_0^t \int_0^t \gamma(t,s_1,s_2,\sigma_1,\sigma_2) a_2(t,\sigma_1,\sigma_2) d\sigma_2 d\sigma_1 \tag{4.51}$$

where

$$F_1(t,s_1,s_2) = F(t,s_1,s_2) - \int_0^t [q(t,s_2,\sigma_1)F(t,s_1,\sigma_2) + q(t,s_1\sigma_2)F(t,\sigma_1,s_2)] d\sigma_2 d\sigma_1$$

$$+ \int_0^t \int_0^t q(t,s_1,\sigma_1)q(t,s_2,\sigma_2)F(t,\sigma_1,\sigma_2) d\sigma_2 d\sigma_1$$

$$\gamma(t,s_1,s_2,\sigma_1,\sigma_2) = \gamma_1(t,s_1,s_2,\sigma_1,\sigma_2) - \int_0^t q(t,s,u)\gamma_1(u,s_2,\sigma_1,\sigma_2) du$$

$$\gamma_1(t,s_1,s_2,\sigma_1,\sigma_2) = - r_2(t,s_1,s_2,\sigma_1,\sigma_2) - q(t,s_2,\sigma_1)q(t,s_1,\sigma_2)$$

$$+ \int_0^t q(t,s_2,u) r_2(t,s_1,u,\sigma_1,\sigma_2) du .$$

Proof From theorem 4.5, it suffices to show that $I + R_1(t)$ is invertible and that 1 is not an eigenvalue of $R(t)$. The invertibility of $I + R_1(t)$ is proven in lemma 4.5. The eigenvalue condition is a consequence of the uniqueness of the best quadratic estimate. For suppose that $c(s_1,s_2) = (R(t)c)(s_1,s_2)$. Then $a_2'(t,s_1,s_2) = a_2(t,s_1,s_2) + c(s_1,s_2)$ would also be a solution of (4.44), and hence, if $a_0'(t)$ and $a_1'(t,s)$ were defined from $a_2'(t,s_1,s_2)$ via (4.43) and (4.41a), $a_0'(t), a_1'(t,s), a_2'(t,s_1,s_2)$ would also satisfy the optimal kernel equations. This contradicts the uniqueness of the best quadratic estimate. The definitions of F_1 and γ follow from

the proof of theorem 4.5. Finally, note that if $\bar{\gamma}$ denotes the integral operator defined by γ, $I + \bar{\gamma}$ is invertible and hence (4.51) has a unique solution.

We have thus reduced the complicated kernel equations (4.40) to a simple Fredholm equation which can be solved by standard methods. Moreover, we can achieve a similar result for the problem of determining the best quadratic correction to a linear filter, which, as previously mentioned, requires solving (4.40c) for a_2 in terms of a_0 and a_1. Again, (4.40c) is an equation of tensor-contraction type for a_2 and the tensor contraction kernels are the same as in (4.44):

$$\begin{aligned} a_2(t,s_1,s_2) = [&Ef(t)h(s_1)h(s_2) - a_0(t)Eh(s_1)h(s_2) - a_1(t,s_1)Eh(s_2) \\ &- a_1(t,s_2)Eh(s_1) - \int_0^t a_1(t,\sigma)Eh(\sigma)h(s_1)h(s_2)d\sigma] \\ &- \int_0^t [r_1(s_2,\sigma)a_2(t,s_1,\sigma) + r_1(s_1\sigma)a_2(t,s_2,\sigma)]d\sigma \\ &- \int_0^t \int_0^{\sigma_1} E\{h(s_1)h(s_2)h(\sigma_1)h(\sigma_2)\} \times a_2(t,\sigma_1,\sigma_2)d\sigma_2 d\sigma_1. \end{aligned}$$

Note that this method does not succeed in solving the optimal kernel equations recursively. Rather, t is fixed throughout and the relevant operators are defined and inverted on $L^2([0,t])$ or $L^2([0,t]^2)$. At a different time t', the entire

process would be repeated. A recursive solution would use $a_2(t,s_1,s_2)$ to construct $a_2(t+dt,s_1,s_2)$.

An important problem is to determine conditions on f, h and the signal process $x(\cdot)$ such that $a_1(t,s)$ and $a_2(t,s_1,s_2)$ are separable, for in this case the filter can be constructed with stochastic differential equations (see theorem 2.3). This has not yet been done and is fairly complicated due to the complex manner in which $Ef(t)$, $Ef(t)h(s)$, etc. combine to produce the kernels F_1 and γ of (4.51), the Fredholm equation for a_2.

CHAPTER 5 NONLINEAR FILTERING PROBLEMS WITH FINITE DIMENSIONAL ESTIMATION ALGEBRAS

Suppose that a signal $x(t)$ is defined by the stochastic differential equation

$$dx(t) = f(x(t))dt + g(x(t))db(t)$$

$$x(0) = x_0 \tag{5.1}$$

and that it is observed via

$$dy(t) = h(x(t))dt + dw(t) \ . \tag{5.2}$$

As usual, $b(t)$ and $w(t)$ are assumed to be independent Brownian motions. The filtering problem associated to (5.1) - (5.2) will be completely solved if the conditional distribution, $P(x(t)\in A|\mathcal{F}_t^y)$, $A \in \{\text{Borel sets of state space}\}$, is known. A basic question is: when can $P(x(t)\in A|\mathcal{F}_t^y)$, as a measure, be characterized by a finite set of statistics propagating recursively in time?

Recent progress on this issue has come from several directions. First, V. Beneš [1] has proved the following result by probabilistic methods. Suppose g = constant, $h(x) = ax + b$, and f is a global solution, (i.e., defined on all of $\mathbb{R}$) of

$$f' + f^2 = ax^2 + bx + c \ . \tag{5.3}$$

Then the conditional distribution of $x(t)$ given $\mathcal{F}_t^y$ has a density which can be expressed in terms of a finite number of statistics

generated by Kalman-Bucy type filtering equations.
There is also a suitable generalization to the case of vector signals and observations. This result covers the case treated by the Kalman-Bucy theorem, (lemma 3.1), but it gives new examples of finite dimensionally computable filters as well. We will refer to the filtering problems treated in Beneš' theorem as the "Beneš examples".

Secondly Brockett and Clark [7], Mitter [35], and Brockett [5 ,6] have suggested that Lie algebraic techniques can be applied to the nonlinear filtering problem. They show how to associate a Lie algebra of operators, the so-called estimation algebra, with filtering models such as (5.1) and (5.2) and how the Lie algebra structure bears upon the filtering solutions. In particular, they suggest that when the estimation algebra is finite dimensional it may be possible to compute conditional densities finite-dimensionally.

In this chapter, we will pursue the implications of Lie algebraic techniques for exact, finite dimensional calculation of conditional densities. The first section will sketch the basic ideas of this theory, especially those that concern generating filter solutions from the estimation algebra structure. As part of this exposition, we will derive by Lie algebraic techniques the conditional density for the problem of estimating a Brownian

motion in white noise. The solution to this problem is, of course, well-known. We derive it here not for the end result but to demonstrate and explore a Lie algebraic technique, called the Wei-Norman method, (Wei and Norman [44]), that establishes the connection between finite dim. estimation algebras and finite dim. filters. Further, a rigorous derivation of a filtering solution directly from the algebraic structure has not appeared in the literature for diffusion signals and so we present one here. As another part of this treatment, we will indicate connections between the Lie algebra strategy and the theory of Lie algebra/Lie group representations on infinite dimensional vector spaces (see also Brockett [6]). This will provide us insights into the behavior of the filtering problems we consider. Also, Lie algebraic theories of estimation are presently very incomplete and we believe representation theory will ultimately offer much to their study. Finally, we briefly develop the Beneš examples from the Lie algebraic viewpoint, a possibility first realized by Mitter, who suggested it to us, after Beneš' results became known.

The remainder of the chapter is devoted to the search for new examples that may be solved Lie algebraically. Section 2 presents a case to which the methods developed in this section do not apply, and it explains why not. Section 3 contains the principal results of the chapter. It first gives necessary conditions for a general class of vector process models to have finite dimensional estimation

algebras, since these are the models that could possibly be solved by the algebraic techniques. By applying this result, we are able to list all possible problems with finite dimensional estimation algebras for scalar process models. We then ask the question: for which of these examples does the Wei-Norman, Lie algebra calculation work? Our results are largely negative. Roughly speaking, they indicate that only for the previously known examples does the calculation work.

5.1 Estimation Algebras and Filtering

Our concern henceforth shall be with the filtering problem stated in (5.1) - (5.2). Suppose that for all t, x(t) has a density q(x,t). Then the conditional distribution of x(t) given $\mathcal{F}_t^y$ has a density

$$\hat{p}(x,t|\mathcal{F}_t^y) = \frac{E_0\{L_t|\mathcal{F}_t^y,\ x(t)=x\}q(x,t)}{E_0\{L_t|\mathcal{F}_t^y\}} . \tag{5.3}$$

This is easily derived from the Kallianpur-Striebel formula. Call the numerator of (5.3) $p(x,t|\mathcal{F}_t^y)$. $p(x,t|\mathcal{F}_t^y)$ captures the x-dependence of $\hat{p}(x,t|\mathcal{F}_t^y)$, that is, it equals $\hat{p}(x,t,|\mathcal{F}_t^y)$ up to a random normalization factor, and it is, therefore, called the unnormalized conditional density. $p(x,t|\mathcal{F}_t^y)$ is easier to work with tnan $\hat{p}(x,t|\mathcal{F}_t^y)$.

Let $L^* = \frac{1}{2}\frac{\partial^2}{\partial x^2} g^2(x) - \frac{\partial}{\partial x} f(x)$ be the forward generator of x(t). Under appropriate regularity conditions, Zaka[[illegible],

Pardoux [40]), the unnormalized conditional density $p(x,t|F_t^y)$ defined in (5.3) satisfies the stochastic partial differential equation

$$dp(x,t) = L^*p(x,t)dt + h(x)p(x,t)dy \qquad (5.4)$$
$$p(x,0) = p_0(x) = \text{initial density of } x(0).$$

The Stratonovich and white noise forms of (5.4) are, respectively,

$$\bar{d}p(x,t) = [L^* - \tfrac{1}{2}h^2(x)]p(x,t)dt + h(x)\, p(x,t)\bar{d}y \qquad (5.5)$$

and

$$\frac{\partial p(x,t)}{\partial t} = [L^* - \tfrac{1}{2}h^2(x)]p(x,t) + h(x)p(x,t)\dot{y}(t). \qquad (5.6)$$

In (5.5), $\bar{d}$ denotes the Stratonovich differential. (5.6) is a formal expression because $\dot{y}(t)$ does not exist except in a generalized sense but it is useful in calculations. (Note: As in (5.4) - (5.6), the $y(\cdot)$ dependence of $p(x,t|F_t^y)$ will often be suppressed for notational convenience.) All or any of these equations will be referred to as Zakai's equation, (Zakai [47]). For the Lie algebraic theory, it is necessary to work with (5.6), since manipulations involving $\dot{y}(t)$ obey ordinary, rather than Ito calculus, and the Lie algebra results to be adapted were developed for deterministic problems with ordinary calculus.

The precise question that we will study here may now be stated. When can $p(x,t)$ be characterized by a finite number of statistics propagating in time; in other words, when does $p(x,t)$ evolve on a finite dimensional manifold? The new approach to filtering that we deal with here is to learn about $p(x,t)$ by applying lie algebra/Lie group methods to Zakai's equation. Let

$L_0 = L^* - \frac{1}{2} h^2(x)$ and let

$\{L_0,h\}_{LA}$ denote the lie algebra of operations generated by L_0 and h using the bracket operation [A,B] = AB-BA. After Brockett [6], we call $\{L_0,h\}_{LA}$ the estimation algebra.

The fundamental idea is that $\{L_0,h\}_{LA}$ carries information about the infinitesimal behavior of p(x,t). In particular, if p(x,t) evolves on a finite dimensional manifold, then $\{L_0,h\}_{LA}$ ought to be finite dimensional as a consequence of the interplay between Lie groups and Lie algebras, (Brockett [5]). Thus, we can search for finite-dimensionally solvable filtering problems by looking for examples in which $\dim\{L_0,h\}_{LA} < \infty$. This will become clearer in the subsequent discussion.

Example 1: (Brockett and Clark [7], Mitter [35]).

Consider the simplest case covered by the Kalman-Bucy theorem (see lemma 3.1).

$$\begin{aligned} x(t) &= x_0 + b(t) \\ dy(t) &= x(t)dt + dw(t) \end{aligned} \tag{5.8}$$

where x_0 is a random variable independent of the processes b(t) and w(t). For (5.8), Zakai's equation is

$$\begin{aligned} &\frac{\partial}{\partial t} p(x,t) = (\frac{1}{2} \frac{\partial^2}{\partial x^2} - \frac{1}{2}x^2)p(x,t) + \dot{y}(t)xp(x,t) \\ &p(x,0) = \text{density of } x_0 . \end{aligned} \tag{5.9}$$

The corresponding estimation algebra is easily seen to be

$$\mathcal{A} \equiv \text{Span } \{\frac{1}{2}\frac{d^2}{dx^2} - \frac{1}{2}x^2, x, \frac{d}{dx}, I\} \quad .$$

Define $A_0 = \frac{1}{2}\frac{d^2}{dx^2} - \frac{1}{2}x^2$. The commutation relationsof $\mathcal{A}$ are

$$[A_0, x] = \frac{d}{dx}$$

$$[A_0, \frac{d}{dx}] = x \qquad (5.10)$$

$$[\frac{d}{dx}, x] = I \quad .$$

We remark that $\mathcal{A}$ is solvable.*

Estimation Algebras and Solutions

In exploring the interaction between finite dimensional estimation algebras and finite dimensionally computable $p(x,t)$ we must first confront the question: Given a finite-dim. estimation algebra how does one integrate it to get a solution of (5.6), i.e., how does one determine $p(x,t)$ from the algebra structure? This problem relates naturally to the theory of integrating representations of Lie algebras of unbounded operators on a Hilbert space to a corresponding representation of a Lie group. We present some

---

*A lie algebra G is solvable if the series of ideals $G_1 = [G,G], \cdots, G_n = [G_{n-1}, G_{n-1}], \cdots$ terminates at 0 for some $n < \infty$.
$I \subset G$ is an ideal if $[I,G] \quad I$.

ideas of this theory to motivate the main calculational method, and because, as mentioned above, we find it a useful source of concepts for thinking about estimation algebras.

Let G be a finite-dim. lie algebra and G its associated (simply connected) group. Let H be a complex Hilbert space.

Definition 5.1

A representation T of G on H is a map T from G onto a set of linear operators on H with a common, dense, invariant domain D such that $[T(x),T(y)] = T([x,y])$ for all $x,y \in G$. Likewise, a representation τ of G on H is a map $\tau: G \to L(H) =$ bounded linear operators on H such that $\tau(g_1 g_2) = \tau(g_1)\tau(g_2)$ for $g_1, g_2 \in G$.

The problem of integrating a Lie algebra representation to a group representation is as follows. Given a representation T of G on H, when does there exist a group representation τ of G such that $\tau(e^{tx}) = e^{tT(x)}$ $\forall$ $x \in G$? Here $e^{tT(x)}$ is a group generated by $T(x)$ in the sense that $\frac{d}{dt} e^{tT(x)}\phi = T(x)\phi$ $\forall$ $\phi \in D$. Suppose that $x_1, \cdots, x_d$ is a basis for G and that we have groups $e^{tT(x_i)}, i=1,\ldots,d$. A method for constructing τ locally is to define

$$\tau(e^{t_1 x_1} \cdots e^{t_d x_d}) = e^{t_1 T(x_1)} \cdots e^{t_d T(x_d)} . \tag{5.11}$$

Formally, this can be made to work, if the operator identity

$$e^{tX_j} X_i = \sum_{n=0}^{\infty} \frac{t^n}{n!} [adX_j]^n X_i e^{tX_j} \quad * \qquad (I)$$

holds for $X_j = T(x_j)$, $1 \leq j, j \leq d$. (Flato, et al. [13]). (I) will be of chief importance.

The procedure recalls the Wei-Norman [44] technique for solving differential equations. Let us develop this formally for the above situation.

Suppose that in H we want to solve the evolution equation

$$\frac{dp}{dt} = X_1 p + u(t) X_2 P \qquad (5.12)$$

which is similar to Zakai's equation. We try a solution in $T(G)$;

$$p(t) = e^{g_1(t)X_1} \cdots e^{g_d(t)X_d} p(0) \quad . \qquad (5.13)$$

For this $p(t)$,

$$\frac{dp}{dt} = \dot{g}_1(t) X_1 p$$
$$+ \dot{g}_2(t) e^{g_1(t)X_1} X_2 e^{g_2(t)X_2} \cdots e^{g_d(t)X_d} p(0) + \cdots$$
$$+ \dot{g}_d(t) e^{g_1(t)X_1} \cdots X_d e^{g_d(t)X_d} p(0) \quad . \qquad (5.14)$$

From (4), for $1 \leq i, j \leq d$

---

*$[adA]B = [A,B]$.

$$e^{tX_j}X_i = \sum_{m=1}^{d} c_m^{i,j}(t)X_m e^{tX_j}$$

and applying this repeatedly in (7),

$$\frac{dp}{dt} = F_1(g(t),\dot{g}(t))X_1 p + \cdots + F_d(g(t),\dot{g}(t))X_d P \qquad (5.15)$$

for some non-linear functions F_i of $g(t) = (g_1(t),\cdots,g_d(t))$ and $\dot{g}(t)$. For $p(t)$ to solve (5), $F_1(g(t),g(t)) = 1$, $F_2(g(t),g(t)) = u(t)$ and $F_j(g(t), \dot{g}(t)) = 0$ for $j > 2$. Solving this set of equations (locally in t) for $g(t)$ gives a local solution of (5).

We will use this method to solve a filtering problem, but first we present a few more remarks on Lie algebra/Lie group representation theory. The heuristic ideas contained in (5.11) and (I) have been worked into a rigorous theory by Nelson [37] and Flato, et al. [13] for the situation in which G is represented by skew-symmetric operators and τ is required to be unitary (i.e., to take values in the space of unitary operators). Their results involve heavy use of the notion of analytic vectors (Nelson [37]); see appendix 2 for a definition. In their theory τ will exist if the algebra domain D contains a dense, invariant set of vectors analytic for each element of a basis for $T(G)$. Conversely, if τ exists there is a common, dense set of analytic vectors for the whole lie algebra (see Flato, et al. [13]). Further implications of the theory are revealed in the following examples, which play a role in

the later discussion.

Example 2 Let $\Lambda_1 = \{-ix, \frac{d}{dx}, -i\}$ and specify its domain as

$$S = \text{rapidly decreasing functions}$$

$$= \{\phi \in C^\infty(\mathbb{R}) \mid \sup_x |x^\beta \partial^\alpha x \phi(x)| < \infty \ \forall \ \alpha, \beta \in Z^+\}$$

Λ_1 is then a representation of the so-called Heisenberg algebra on $L^2(\mathbb{R})$ and it does generate a Lie group on $L^2(\mathbb{R})$. It is not important here to present this group. However, it is interesting to construct a domain of analytic vectors for Λ_1 because this involves the second order operator $A_0 = \frac{1}{2}(\frac{d}{dx})^2 + \frac{1}{2}(-ix)^2 = \frac{1}{2}\frac{d^2}{dx^2} - \frac{1}{2}x^2$, which arose in the estimation algebra Λ of example 1. Indeed, it turns out that A_0 on $L^2(\mathbb{R})$ has a discrete spectrum $\{\lambda_n\}_{n=1}^\infty$, $\lambda_n < 0$, $\lim_{n\to\infty} \lambda_n = -\infty$. Let $\{\psi_n\}_{n=1}^\infty$ be the corresponding eigenvectors, and define

$$D' = \{ \sum_1^N \alpha_n \psi_n \mid N < \infty \}$$

D' is a dense invariant set of analytic vectors for A_0 and Λ_2.

It is easily seen that D' is also a dense set of analytic vectors for Λ of example 1.

Example 3 Let $\Lambda_2 = \{-ix, \frac{d}{dx}, -i\}$ be a representation of the Heisenberg algebra on $L^2(\mathbb{R}^+)$ with domain $C_0^\infty(\mathbb{R}^+) \equiv \{\phi \in C^\infty(\mathbb{R}^+) \mid \phi$ has compact support in $\mathbb{R}^+\}$. In this case, a unitary representation

generated by A_2 does not exist. The reason is that $i\frac{d}{dx}$ will not be essentially self-adjoint on $C_0^\infty(\mathbb{R}^+)$ (see Appendix 2), and, by a theorem of Nelson [37], this precludes τ. Nelson [37] also shows that $\frac{1}{2}\frac{d^2}{dx^2} - \frac{1}{2}x^2 = \frac{1}{2}(\frac{d}{dx})^2 + \frac{1}{2}(-ix)^2$ will not be essentially self-adjoint on $C^\infty(\mathbb{R}^+)$ and that $C_0^\infty(\mathbb{R}^+)$ will not contain a dense, invariant set of analytic vectors either. We shall observe analogous behavior for the filtering problem studied in section 2.

We will now adapt the Wei-Norman method of (5.12) - (5.15) to the solution of (5.9) in example 1. The first step is to associate evolutions e^{tX_i} to the elements X_i of the estimation algebra Λ. However, for Λ, the situation is considerably more complicated than in the theory of representations by skew-symmetric operators. It will no longer always be possible to generate groups with the elements of Λ, or to insure that the evolutions are bounded. Nevertheless, we proceed with the most natural definitions. Define $\{e^{tA_0} | t\geq 0\}$ to be the semigroup associated to A_0 by solving in $L^2(\mathbb{R})$ the equation

$$\dot{u}(t) = A_0 u(t) \qquad u(0) = \phi \in L^2(\mathbb{R}) \quad .$$

It is well-known that

$$(e^{tA_0}\phi)(x) = \int_{-\infty}^{\infty} G(x,y,t)\phi(y)dy \qquad t \geq 0 \tag{5.16}$$

$$G(x,y,t) = (2\pi\sinh t)^{-1/2} \exp[-\tfrac{1}{2}(\coth t)(x^2+y^2) + xy/\sinh t]$$

Similary, define

$$(e^{tx}\phi)(x) = e^{tx}\phi(x)^{*} \qquad t \in \mathbb{R}$$

$$(e^{t\frac{d}{dx}}\phi)(x) = \phi(x+t) \qquad t \in \mathbb{R} .$$

As in (5.13) let us try to solve (5.9) by the expression

$$p(x,t) = [e^{g_1(t)A_0}\, e^{g_2(t)x}\, e^{g_3(t)\frac{d}{dx}}\, e^{g_4(t)}\, p_0]x \qquad (5.17)$$

where the $g_i(t)$ functions are to be determined. One may certainly raise objections to (5.17) -- the operator $e^{g_2(t)x}$ is unbounded, $g_1(t)$ cannot take negative values -- but these will be cleared up as we go along. The Wei-Norman method will allow the values of $g_i(t)$ to be calculated if (I) holds for the elements of A. This requires, for example, that

$$e^{tA_0}x\phi = (\cosh t)x\, e^{tA_0}\phi + (\sinh t)\frac{d}{dx}\, e^{tA_0}\phi \qquad (5.18)$$

$$e^{tA_0}\frac{d}{dx}\phi = (\sinh t)x\, e^{tA_0}\phi + (\cosh t)\frac{d}{dt}\, e^{tA_0}\phi \qquad (5.19)$$

and

$$e^{tx}\frac{d}{dx}\phi = \frac{d}{dx}\, e^{tx}\phi - t\, e^{tx}\phi \quad . \qquad (5.20)$$

The right-hand-sides of (5.18) and (5.19) are derived from (I) by using the identities

$$[\mathrm{ad}\, A_j]^{2n}x = x$$

*This solves $u(t) = xu(t)$ $u(0) = \phi$, but not in $L^2(\mathbb{R})$.

$$[\text{ad } A_0]^{2n+1} x = \frac{d}{dx}$$

which follow easily from the commutation relations (5.10).

Lemma 5.1

(i) (5.20) holds for every $\phi \in C^1(\mathbb{R})$

(ii) Let $V = \{\phi(x) = x^\beta e^{\alpha x}\psi(x) \mid \beta,\alpha\in\mathbb{R}, \psi \in L^1(\mathbb{R})\}$. For every $\phi \in V$

$$(e^{tA_0}\phi)(x) = \int_{\mathbb{R}} G(x,y,t)\phi(y)dy$$

exists and is infinitely differentiable on $\{(x,t) \mid t\geq 0\}$. Further

$$\frac{\partial^{m+n}}{\partial t^m \partial x^n}(e^{tA_0}\phi)(x) = \int_{-\infty}^{\infty} \frac{\partial^{m+n}}{\partial t^m \partial x^n} G(x,y,t)\phi(y)dy \qquad (5.21)$$

and for $t > 0$, $\frac{\partial}{\partial t}(e^{tA_0}\phi)(x) = (A_0 e^{tA_0}\phi)(x)$ (5.22)

(iii) (5.18) is true for every $\phi \in V$, (5.19) for every $\phi \in V$ such that $\phi' \in V$.

Remark If $p_0 \in L^1(\mathbb{R})$ then $e^{g_2(t)x}\, e^{g_3(t)\frac{d}{dx}}\, e^{g_4(t)}\, p_0 \in V$ and hence $p(x,t)$, as given by the product of evolutions (5.17), is a well-defined function in $L^1(\mathbb{R})$.

Proof (i) is an elementary calculation.

(ii) can be derived easily using the explicit form of $G(x,y,t)$. The principal observation to make in doing the calculations is that for $t > 0$, $G(x,y,t)$ will decay like $e^{-\alpha y^2}$ as $y \to \infty$ for some positive α.

(iii) The following calculation proves (5.18)

$$(e^{tA_0}x\phi)(z) = \int_{-\infty}^{\infty} G(z,y,t)y\phi(y)dy$$

$$= \int_{-\infty}^{\infty} [\frac{d}{dy} G(x,y,t)|_{x=z} + (\coth t)x\ G(x,y,t)|_{x=z}]$$

$$\sinh t\ \phi(y)dy$$

$$= (\sinh t \frac{d}{dx} e^{tA_0}\phi)(z) + (\cosh t)(xe^{tA_0}\phi)(z).$$

To get, (5.19) integrate by parts

$$(e^{tA_0} \frac{d}{dx}\phi)(z) = \int G(z,y,t)\frac{d}{dy}\phi(y)dy = -\int [\frac{d}{dy} G(z,y,t)]\phi(y)dy$$

$$= (\coth t)(e^{tA_0}x\phi)(z) - \frac{1}{\sinh t}(xe^{tA_0}\phi)(z)$$

$$= \sinh t\ (xe^{tA_0}\phi)(z) + \cosh t\ (\frac{d}{dx} e^{tA_0}\phi)(z).$$

To obtain the last equality, we used (5.18).

Let $p(x,t)$ be as in (5.17). We will now solve for the functions $g_i(t)$ $i = 1,2,3,4$. From (5.17)

$$\frac{\partial p}{\partial t} = \dot{g}_1(t)A_0p + \dot{g}_2(t)\ e^{g_1(t)A_0}x\ e^{g_2(t)x}\ e^{g_3(t)\frac{d}{dx}}\ e^{g_4(t)}\ p_0$$

$$+ \dot{g}_3(t)\ e^{g_1(t)A_0}\ e^{g_2(t)x\frac{d}{dx}}\ e^{g_3(t)\frac{d}{dx}}\ e^{g_4(t)}\ p_0$$

$$+ \dot{g}_4(t)p \qquad (5.23)$$

By applying (5.13) - (5.20), one derives

$$\frac{\partial p}{\partial t} = \dot{g}_1(t)A_0 p(x,t) + F_2(g(t),\dot{g}(t))xp(x,t) + F_3(g(t),\dot{g}(t))\frac{d}{dx}p(x,t)$$

$$+ F_4(g(t),\dot{g}(t))p(x,t) \tag{5.24}$$

with F_i's given below in (5.25). But if $p(x,t)$ is to solve (5.9) with $p(x,0) = p_0(x)$ we must require

$$\dot{g}_1(t) = 1$$

$$\dot{y}(t) = F_2(g(t),\dot{g}(t)) = \dot{g}_2(t)\cosh g_1(t) + \dot{g}_3(t)\sinh g_1(t) \tag{5.25}$$

$$0 = F_3(g(t),\dot{g}(t)) = \dot{g}_2(t)\sinh g_1(t) + \dot{g}_3(t)\cosh g_1(t)$$

$$0 = F_4(g(t),\dot{g}(t)) = \dot{g}_4(t) - \dot{g}_3(t)\, g_2(t)$$

$$g_i(t) = 0 \quad i = 1,\cdots,4$$

(5.25) may be easily solved. The result, written in terms of the $dy(t)$ notation rather than $\dot{y}(t)$, is

$$g_2(t) = \int_0^t \cosh(s)\, dy(s)$$

$$g_3(t) = -\int_0^t \sinh(s)\, dy(s) \tag{5.26}$$

$$g_4(t) = \int_0^t (\sinh s)(\cosh s)ds - \int_0^t g_2(s)(\sinh s)dy(s).$$

Finally, by substituting these expressions in (5.17) and using the explicit form of the kernel $G(x,y,t)$ of e^{tA_0}, we derive

$$p(x,t) = \int_{-\infty}^{\infty} k(z,t)\, e^{-1/2\, p^{-1}(t)[x-m(t)]^2} p_0(z)dz \qquad (5.27)$$

$$p(t) = \tanh t$$

$$m(t) = \frac{z}{\cosh t} + \int_0^t \frac{\sinh s}{\cosh t}\, dy(s) \quad .$$

Let $p(x,t;z)$ denote the integrand of (5.27) exclusive of $p_0(z)$. $p(x,t;z)$ may be interpreted as the unnormalized conditional density of $x(t)$ for the process $x(t)$ starting at $x(0) = z$. It is clear that the normalized version of $p(x,t;z)$ is a Gaussian density with mean $m(t)$ and covariance $p(t)$, and this agrees with the Kalman-Bucy solution of the problem. Though these calculations used the $\dot{y}(t)$ formalism, they can be carried out, with some added computational complexity, using the rigorous Ito calculus. Therefore, we have rederived the Kalman filter.

Remark In presenting the Wei-Norman technique, it was indicated that in general it only gives solutions local in t. However, theorem 5.1 provides a solution for all $t \geq 0$. This happens because the estimation algebra A is solvable (see example 1). Wei and Norman [44] show that for solvable matrix Lie algebras, if the operators $e^{g_i(t)X_i}$ are placed in the correct order in (5.13), global solutions can be found. Without further elaboration, we observe that this result extends to the present case and motivates putting $e^{g_0(t)A_0}$ first in (5.17).

The Beneš Example

The Lie algebra strategy is also able to recover the theorem of V. Beneš quoted above. Consider the scalar case

$$\begin{aligned} dx(t) &= f(x(t))dt + db(t) \\ dy(t) &= x(t)dt + dw(t) \\ f' + f^2 &= ax^2 + bx + c \quad . \end{aligned} \tag{5.29}$$

The Zakai equation is then

$$\frac{\partial p}{\partial t} = \{\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{\partial}{\partial x} f(x) - \frac{1}{2}x^2\}p(x,t) + \dot{y}(t)xP \quad . \tag{5.30}$$

The estimation algebra $\tilde{A} = \{\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{\partial}{\partial x} f(x) - \frac{1}{2}x^2, x\}_{LA}$ is again finite dimensional. Indeed, if $L_0 \equiv \frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{\partial}{\partial x} f(x) - \frac{1}{2}x^2$,

$$\tilde{A} = \text{Span}\{L_0, x, \frac{d}{dx} - f, I\}$$

and its commutation relations are

$$\begin{aligned} [L_0, x] &= \frac{d}{dx} - f \\ [L_0, \frac{d}{dx} - f] &= (a+1)x + \frac{b}{2} \\ [\frac{d}{dx} - f, x] &= I \quad . \end{aligned} \tag{5.31}$$

These calculations are valid without restriction on the coefficients a, b, and c. However, for general a, b, c, (5.29) may not have a global solution, in which case (5.31) is to be interpreted on functions

whose domains are contained in the region where f is defined. We will explore when (5.29) has global solutions in section 5.3.

We could now try to solve the Zakai equation (5.30) by

$$p(x,t) = (e^{g_1(t)L_0}\, e^{g_2(t)x}\, e^{g_3(t)\frac{d}{dx}} - f\, e^{g_4(t)}\, p_0)(x,t) .$$

However, it is simpler to first rewrite Zakai's equation via a simple transformation that compares to the gauge transformations of quantum physics. Let $z \in \mathbb{R}$ and define

$$F(x) = \int_z^x f(s)ds \quad .$$

If $q(x,t) = e^{-F(x)}p(x,t)$, a substitution in (5.30) demonstrates that

$$\frac{\partial q}{\partial t} = [\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}[(a+1)x^2+bx+c]]q + \dot{y}(t)xq \qquad (5.32)$$

(5.32) is similar to the Zakai equation for the Brownian signal example. Although (5.32) is not the Zakai equation for a filtering problem, let us define its Lie algebra of operators as $\tilde{\Lambda} = \{\tilde{L}_0, x\}_{LA}$ $\tilde{L}_0 = \frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}[(a+1)x^2+bx+c]$. $\tilde{\Lambda}$ is isomorphic to $\tilde{\Lambda}$ by the ismorphism $\tilde{L}_0 \to L_0$, $x \to x$, $\frac{d}{dx} - f \to \frac{d}{dx}$, $I \to I$. It is trivial to observe that

$$\tilde{\Lambda} = \text{Span}\{\tilde{L}_0 = \frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{(a+1)}{2}x^2,\ x,\ \frac{d}{dx},\ I\} .$$

Suppose that f is globally defined so that, in order to find $p(x,t)$, we seek a solution $q(x,t)$ of (5.32) that is defined for all $x \in \mathbb{R}$. Suppose further that $(a+1) > 0$. We can then solve (5.32) by the method

$$q(x,t) = (e^{g_1(t)\tilde{A}_0} e^{g_2(t)x} e^{g_3(t)\frac{d}{dx}} e^{g_4(t)} q_0)(x) .$$

The $g_i(t)$ functions will be slightly different from those calculated for example 1 because of the different commutation relations between $\tilde{A}_0, x, \frac{d}{dx}$ and I. The resulting solution $p(x,t) = e^{F(x)}q(x,t)$ is precisely that obtained by Beneš. Whether the same method can be made to work for the case in which f is not globally defined is an issue we will take up in the remaining sections.

5.2 Estimation of Absorbed Brownian Motion

This section counterpoints the successes of 5.1 by providing an example in which the estimation algebra has finite dimension, but in which the Wei-Norman calculation does not work. The signal in this example is a Brownian motion absorbed at a boundary, and the associated filtering problem is very similar, operator-wise, to the problems treated in 5.1. But it turns out that the Wei-Norman method cannot proceed because the identity (I) fails for the estimation algebra operators. The phenemenology of this failure is discussed in hope of characterizing the estimation algebras for which the technique does or does not work. For the examples at hand,

crucial information is revealed by the algebra's behavior on the eigenvectors of the partial differential operator $L_0 = L^* - \frac{1}{2}h^2$ (see 5.6) of Zakai's equation and by the interaction between operators and boundary conditions. The domain structures of the estimation algebras are also more closely identified. An important role is played by the existence or non-existence of a dense invariant domain of analytic vectors associated with the eigenvalues of L_0.

Let $b(t)$ be a Brownian motion, let x_0 be a r.v. with density $p_0(x)$ such that $p_0(x) = 0, x \leq 0$ and such that x_0 is independent of $b(t)$, and let $\tau = \inf\{t | x_0 + b(t) = 0\}$. Consider the problem

$$\begin{aligned} x(t) &= (x_0 + b(t))\, 1_{\{t<\tau\}} \\ dy(t) &= x(t)dt + dw(t) \end{aligned} \tag{5.33}$$

$x(t)$ is an absorbed Brownian motion with random initial value. The distribution will now have two parts; an atom $Q_0(t) = P_a\{x(t)=0\}$, and a measure $Q(A,t) = P_a\{x(t)\in A-\{0\}\}$ (for $A \subset [0,\infty)$) with density $q(x,t)$. Accordingly, the unnormalized conditional density of $x(t)$ will have two parts:

$$P_a(t) = E_0\{1_{\{x(t)=0\}} L_t | F_t^y\}$$

and

$$P(A,t) = E_0\{1_{\{x(t)\in A-\{0\}\}} L_t | F_t^y\}$$

$P(A,t)$ will have a density

$$p(x,t) = q(x,t)\, E_0\{1_{\{t<\tau\}} \exp\Big[\int_0^t (x_0+b(s))dy(s) - \frac{1}{2}\int_0^t (x_0+b(s))^2 ds\Big] \mid \mathcal{F}_t^y,\ x(t) = x\}. \qquad (5.34)$$

Again these statements are all consequences of the Kallianpur-Striebel formula. Assuming differentiability, $p(x,t)$ will satisfy the Zakai equation

$$\frac{\partial p(x,t)}{\partial t} = (\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}x^2)p(x,t) + \dot{y}(t)xp(x,t) \qquad (5.35)$$

$$p(0,t) = 0 \quad t > 0$$

$$p(x,0) = p_0(x) \quad .$$

(5.35) may be derived formally as follows. Let

$$V = \{f \in C(\mathbb{R}^+) \mid f \text{ has finite limit at } +\infty\} .$$

Let $(T(t)f)(x)$ be the semigroup on V generated by absorbed Brownian motion. The generator of $T(t)$ is $\frac{1}{2}\frac{d^2}{dx^2}$, with domain $D = \{f \in C^2 \mid f'' \in V,\ f''(0)=0\}$ (see, e.g., Lamperti [49]). For $f \in D$, define

$$\sigma_t(f) = E_0[f(x(t))L_t \mid \mathcal{F}_t^y] \qquad (5.36)$$
$$= f(0)P_a(t) + \int_{(0,\infty)} f(x)p(x,t)dx \quad .$$

By applying the stochastic differential equations of filtering (Fujisaki, Kallianpur, Kunita [15] and Kunita [25]) one may derive

$$d\sigma_t(f) = \sigma_t(\frac{1}{2}\frac{d^2}{dx^2} f)dt + \sigma_t(xf)dy(t) \qquad (5.37)$$

for $f \in D$. By substituting (5.36) in (5.37) and integrating $\sigma_t(\frac{1}{2}\frac{d^2}{dx^2} f)$ by parts, one finds

$$f(0)dP_a(t) + \int_{(0,\infty)} f(x)\, dp(x,t)dx$$

$$= [-\frac{1}{2} f'(0)p(0,t) + \frac{1}{2} f(0)\frac{\partial}{\partial x} p(x,t)|_{x=0}]dt$$

$$+ \int_{(0,\infty)} f(x)\frac{1}{2}\frac{\partial^2}{\partial x^2} p(x,t)dt\, dx$$

$$+ \int_{(0,\infty)} f(x)xp(x,t)dy(t)dx \quad .$$

Since this must hold for all $f \in D$

$$dP_a(t) = \frac{1}{2}\frac{\partial}{\partial x} p(x,t)|_{x=0}\, dt$$

$$p(0,t) = 0$$

$$dp(x,t) = \frac{1}{2}\frac{\partial}{\partial x^2} p(x,t)dt + xp(x,t)dy(t) \quad . \qquad (5.38)$$

The white noise version of (5.38) is indeed (5.36).

In (5.35), let A, rather than A_0, denote $\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}x^2$, to emphasize that the functions A acts on are now defined only on $\mathbb{R}^+$. The estimation algebra is

$$\mathcal{A}' = \{A,\ x,\ \frac{d}{dx},\ I\}$$

which, but for the fact that the domain of functions is different, is the same as $\mathcal{A}$ in example 1 of section 5.1. Again, one might hope that $p(x,t)$ is solved by

$$p(x,t) = (e^{g_1(t)A}\ e^{g_2(t)x}\ e^{g_3(t)\frac{d}{dx}}\ e^{g_4(t)}\ p_0)(x), \qquad (5.39)$$

once the various operators are properly defined. However, the crucial identity (I) will fail in this case. The problem is the boundary condition $p(0,t) = 0$. Because of this condition, e^{tA} will not be the same as e^{tA_0} of the previous section. Rather, to meet this condition in (5.39), we must require. $(e^{tA}\phi)(0) = 0$ for all relevant ϕ. A simple reflection argument on the kernel $G(x,y,t)$ of e^{tA_0} (see 5.16) yields

$$(e^{tA}\phi)(x) = \int_0^\infty G_1(x,y,t)\phi(y)dy$$

$$G_1(x,y,t) = \sqrt{\frac{2}{\pi \sinh t}}\ \exp[-\tfrac{1}{2}\coth t(x^2+y^2)]\sinh(\frac{xy}{\sinh t}). \qquad (5.40)$$

If we try to prove the analogue of (5.61) with A replacing A_0 we

then get

$$(e^{tA}x\varphi)(z) = \cosh t(xe^{tA}\varphi)(z) + \sinh t(\frac{d}{dx}\, e^{tA}\varphi)(z)$$

$$+ \int_0^\infty \sqrt{\frac{2}{\pi \sinh t}}\, e^{-1/2\, \coth t(z^2+y^2)}\, e^{-zy/\sinh t}\varphi(y)dy.$$

The last term will not be identically zero unless φ is, and hence (I) fails. The Lie algebraic calculation of section 1 to solve Zakai's equation will then not work.

Discussion

In the above calculations, we verified or disproved the crucial identity (I) by using the explicit formulae for e^{tA} and e^{tA_0}. It is desirable to explain the results at a more fundamental, operator-theoretic level, i.e., to understand how the closely related estimation algebras A and A' involve such widely variant behavior.

Our first step is to reprove the identities (5.18) and (5.19) of section 5.1 by much more fundamental methods. This proof will not be quite as strong as that of lemma 5.1 since it will apply only to restricted set of functions. Recall from example 2 of section 5.1 that

D' = {finite linear combinations of eigenvectors}

is a dense, invariant domain for A. Let $\psi \in D'$ be an eigenvector

with eigenvalue λ. From the commutation relations (5.10)

$$\frac{d}{dx}\psi = [A_0, x]\psi = A_0 x\,\psi - x\lambda\psi \tag{5.41}$$

$$x\psi = [A_0, \frac{d}{dx}]\psi = A_0\,\frac{d}{dx}\psi - \frac{d}{dx}\,\lambda\psi\ . \tag{5.42}$$

By adding (5.40) and (5.41) we derive

$$(\frac{d}{dx} + x)\psi = A_0(\frac{d}{dx} + x)\psi - \lambda(\frac{d}{dx} + x)\psi\ .$$

Since D' is invariant under $\frac{d}{dx}$ and x, $(\frac{d}{dx} + x)\psi \in D'$ also. Thus we conclude that $(\frac{d}{dx} + x)\psi$ is an eigenvector of A_0 with eigenvalue $\lambda + 1$. A similar argument shows that $\frac{d}{dx} - x$ is an eigenvector with eigenvalue $\lambda - 1$. The following calculation now proves (5.18) for $\phi = \psi$. Observe first that $e^{tA_0}\psi = e^{\lambda t}\psi$,

$e^{tA_0}(\frac{d}{dx} + x)\psi = e^{t(\lambda+1)}(\frac{d}{dx} + x)\psi$, etc. Then

$$\begin{aligned} e^{tA_0}\, x\,\psi &= e^{tA_0}\tfrac{1}{2}[(\tfrac{d}{dx} + x)\psi - (\tfrac{d}{dx} - x)\psi] \\ &= \tfrac{1}{2}\, e^{t(\lambda+1)}(\tfrac{d}{dx} + x)\psi - \tfrac{1}{2}\, e^{t(\lambda-1)}(\tfrac{d}{dx} - x)\psi \\ &= \cosh t\ x\, e^{t\lambda}\psi + \sinh t\ \frac{d}{dx}\, e^{t\lambda}\psi \\ &= \cosh t\ x\, e^{tA_0}\psi + \sinh t\ \frac{d}{dx}\, e^{tA_0}\psi\ . \end{aligned} \tag{5.43}$$

By linearity, (5.43) extends to all the elements of D'. Thus, the invariance of the domain D', which is also a domain of vectors analytic for A, implies identity (I) after a direct calculation using little more than the commutation relations. The development here is reminiscent of the quantum field theory of the harmonic oscillation or the construction of the free Boson field. This is not accidental, because deep connections between quantum field theory and filtering exist. (This was discovered and treated by Mitter [35].)

These nice domain and eigenvector properties of the Brownian motion signal case do not extend to the absorbed Brownian motion problem, despite the isomorphism between A and A'. The root cause is the boundary condition $p(0,t) = 0$, and the fact that this condition is not invariant under $\frac{d}{dx}$. Indeed, we can see intuitively that

$$e^{tA}x\phi \neq \cosh t \; x \; e^{tA}\phi + \sinh t \; \frac{d}{dx} \; e^{tA}\phi \qquad (5.44)$$

because, for general ϕ $(e^{tA}x)\phi(0) = 0$ and $\sinh t(\frac{d}{dx} \; e^{tA}\phi)(x) \neq 0$.

However, the fact that (I) fails is not apparent directly from the structure of A' because the boundary condition makes no contribution to the definition of A'. We can rectify this situation by more careful attention to the issue of operator domains. It is useful to think of A' as a representation of a Lie algebra on a function space V, and in this discussion it will suffice to

set $V = L^2(S)$, S = state space. Recall that a Lie algebra representation required not only an algebra of operators, but also a dense, common, invariant domain on which to define them, because an unbounded operator is not fully specified until its domain is given. Thus it is actually inadequate to discuss estimation algebras without considering domains, and thus we attempt the more rigorous formulation

Definition 5.2 Let $D \subset V$ be a common, dense, invariant domain of L_0 and h. Then $\{L_0,h\}_{LA,D}$ denotes the lie algebra of operators generated by L_0/D and h/D and defined on the domain D.

Remark Domain invariance insures that all brackets $[A,B]$ of elements of the lie algebra are again well-defined operators on D.

What is the correct domain D to associate to Λ' when trying to solve (5.35) by the Wei-Norman product series (5.39)? Up to now, the discussion of operators in Λ' has been formal since we did not specify domains. However, we did find that A, in conjunction with the boundary condition $p(0,t) = 0$, gives rise to the semi-group e^{tA} defined in (5.40). The proper domain D' for A is then that for which $A|D'$ is the infinitesimal generator of e^{tA}. For clarity, denote this infinitesmal generator by $\tilde{A}$. It is easily seen that

$$D(\tilde{A}) = \{\phi \in L^2(\mathbb{R}^+) \mid (\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}x^2)\phi \in L^2(\mathbb{R}^+), \phi(0) = 0\}$$

Henceforth, we will discuss $\tilde{A}$ instead of the less well-defined A. Eigenvectors of $\tilde{A}$ will again be important, and domain considerations enter into their definition; ψ is an eigenvector of $\tilde{A}$ if $(\frac{1}{2}\frac{d^2}{dx^2} - \frac{1}{2}x^2)\psi = \psi$ for some λ and if $\psi(0) = 0$.

Now, in analogy with Lie algebra representation theory, we want the domain D of A' to be such that $\tilde{A}|D$ generates e^{tA} also, (in the sense that e^{tA} is the unique semigroup s.t. $\lim_{t\to 0} \frac{1}{t}[e^{tA}\phi - \phi] = \tilde{A}\phi$, $\forall\, \phi \in D$). At the very least this requires that $D \subset D(\tilde{A})$. Otherwise $\tilde{A}|D$ will generate a different semigroup or will have extensions generating different semigroups. For arbitrary D, of course, it may not be possible to associate any semigroup to $\tilde{A}|D$.

However, the next theorem will show that the requirement $D \subset D(\tilde{A})$ is also problematic and it will lead us to a deeper characterization of why the Wei-Norman method fails.

Theorem 5.2 Let $D \subset D(\tilde{A})$ be a dense (in $L^2(\mathbb{R}^+)$) invariant domain for A'. Then D does not contain any eigenvectors of $\tilde{A}$ nor does it contain a dense, invariant domain of analytic vectors.

Remark It can be shown that $\tilde{A}|D$ has many different self-adjoint, negative extensions. Thus there will exist many other semigroups $U(t)$ such that

$$\lim_{t\downarrow 0} \frac{1}{t} [U(t)\phi - \phi] = A\phi, \quad \forall \phi \in D$$

We conclude that D does not have the structure necessary for integrating the elements of Λ'.

Proof If ψ is an eigenvector of $\tilde{A}$ it is clear that $\psi(0) = 0$ and $\psi'(0) \neq 0$, for otherwise the unique solution of $\frac{1}{2}\psi'' - \frac{1}{2}x^2\psi - \lambda\psi = 0$ (λ = eigenvalue) is $\psi \equiv 0$. However, if $D \subset D(\tilde{A})$ is invariant under Λ', it is at least invariant under $\frac{d}{dx}$. Thus if $\phi \in D$, $\phi^{(n)} \in D \subset D(\tilde{A})$ $\forall n$ which implies $\phi^{(n)}(0) = 0$, $\forall n$. Thus

$$D \subset D(A) \cap \{\phi | \in C^\infty(\mathbb{R}^+),\ \phi^{(n)}(0) = 0,\ \forall n\}\ .$$

It is immediately clear that D contains no eigenvectors of $\tilde{A}$. Consider a representation of the Heisenberg algebra $\Lambda'' = \{-ix, \frac{d}{dx}, -i\}$ on D. As in example 3 of section 5.1, $i\frac{d}{dx}$ is not essentially self-adjoint on D and hence, by the theorem of Nelson [37] (see Appendix 2) Λ'' on D does not integrate to a unitary group and hence does not possess a dense invariant domain of analytic vectors in D. An analytic vector for ix is an analytic vector for x and vice versa, so D does not contain such a domain for Λ' either.

In short, the eigenvectors of $\tilde{A}$, which in the Brownian

signal example were the source of a dense, invariant domain of analytic vectors, are no longer invariant under the estimation algebra. Thus if ψ is an eigenvector of $\tilde{A}$, $\tilde{A}\psi = \lambda\psi$, it is no longer true that $(\frac{d}{dx} + x)\psi$ and $(\frac{d}{dx} - x)\psi$ are eigenvectors of $\tilde{A}$ and hence the proof of (I) on eigenvectors by commutation relations also fails for the absorbed Brownian motion case.

In summary, when a dense invariant domain of analytic vectors, in these cases provided by the eigenvectors of A or A_0, fails to exist in the domain of the estimation algebra, the Lie algebraic method of solving Zakai's equation does not work. We shall see this same behavior repeated in examples presented in the next section. It is our conjecture that the existence of analytic vectors for the domain of an estimation algebra will be a necessary condition that a filtering problem with finite-dim. estimation algebra also have a finite dimensionally computable conditional density. Further work on this has not been done.

5.3 Finite Dimensional Estimation Algebras.

In this section we seek to identify those filtering problems that possess finite-dimensional estimation algebras. We will restrict our attention to the class of models

$$\begin{aligned} dx(t) &= f(x(t))dt + G\, db(t) \\ dy(t) &= h(x(t))dt + dw(t). \end{aligned} \tag{5.45}$$

and we will assume that $x(t)$ is an $\mathbb{R}^n$-valued process, $y(t)$ is $\mathbb{R}^p$-valued, $b(t)$ is an m-dimensional Brownian motion, $m \geq n$ and G is a constant $n \times m$ matrix of full rank. Additionally, we will suppose that $x(t)$ evolves in an open, connected set $U \subset \mathbb{R}^n$, and that $f, h \in C^\infty(U)$. As the estimation algebra domain, we will always take

$$C_0^\infty(U) = \{\phi \in C^\infty(U) \mid \operatorname{supp}\phi \text{ is compact}\},$$

a choice avoiding boundary conditions, but imposing no loss of generality to estimation algebra calculations, since the algebra operators should be defined on a sufficiently well-behaved and complete domain. The first result will present a necessary condition that (5.45) has a finite-dim. estimation algebra. We then use this condition to list all possible finite-dim. examples in the scalar version of (5.45), $n = m = p = 1$. Finally we discuss in which of the scalar possibilities, Zakai's equation can be solved by the method of section 5.1.

Conditions for Finite Dimensionality

The Zakai equation for (5.45) is

$$\frac{\partial p(x,t)}{\partial t} = L_0 p(x,t) + \left(\sum_{i=1}^{p} h_i(x) y_i(t)\right) p(x,t)$$

$$L_0 = \frac{1}{2} \sum_{i,j=1}^{n} A_{ij} \frac{\partial^2}{\partial x_i \partial x_j} - \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f_i(x) - \frac{1}{2} \sum_{i=1}^{p} h_i^2(x) . \qquad (5.46)$$

$$A = GG^T .$$

Consequently, the estimation algebra is defined to be

$$\Lambda = \{L_0, h_1, \cdots, h_p\}_{LA}, C_0^\infty(U) \quad .$$

The following lemma, which incorporates an important Lie algebra calculation, is needed in the first theorem.

Lemma 5.2 Let $g \in C^\infty(U)$. Then $\forall k$

$$[\text{ad } L_0]^k g = \sum_{\ell_1, \cdots, \ell_k = 1} D^k g(A_{\ell_1}, \cdots, A_{\ell_k}) \frac{\partial^k}{\partial x_{\ell_1} \cdots \partial x_{\ell_k}} \qquad (5.47)$$

$$+ \text{ terms with lower order differential operators}$$

In (5.47) $D^k g(\cdots)$ denotes the k^{th} differential of g considered as a symmetric, k-linear function, and A_ℓ denotes the ℓ^{th} column of A.

Proof For k = 1, a direct computation will show

$$[L_0, g]\phi = \sum_{i,j=1}^{n} \frac{\partial g}{\partial x_i} A_{ij} \frac{\partial \phi}{\partial x_j} + [\frac{1}{2} \text{tr } (A \cdot \bar{D}^2 g) - \nabla g \cdot f]$$

for $\phi \in C_0^\infty(U)$. In the last term,

$$\bar{D}^2 g = [D^2 g(e_\ell, e_k)] \ ,$$

$$e_k = k^{th} \text{ standard basis vector}$$

and

$$\nabla g = (\frac{\partial g}{\partial x_1}, \cdots, \frac{\partial g}{\partial x_n}).$$

For $k > 1$, the proof proceeds by induction. The details will not be presented.

The next theorem gives the necessary condition for the model (5.45).

Theorem 5.3 If $\dim \Delta < \infty$, $h_1(x),\cdots,h_p(x)$ must be polynomials of degree ≤ 2. More generally, if $g(x) \in C^\infty(U)$ is in Δ, $g(x)$ must be a polynomial of degree ≤ 2.

Proof Fix $g(x)$. The sequence $\{[\text{ad } L_0]^k g\}_{k=0}^\infty$ is contained in Δ and hence cannot have operators of arbitrarily high order. Because of lemma 5.4, this implies that, for some k,

$$D^k g(A_{\ell_1},\cdots,A_{\ell_k}) = 0$$

for all $\ell_1,\cdots,\ell_k$, $1 \leq \ell_1,\cdots,\ell_k \leq n$. Since A is non-singular, this means that

$$\frac{\partial^k}{\partial x_{\ell_1} \cdots \partial x_{\ell_n}} g = 0$$

for all $\ell_1,\cdots,\ell_k$ $1 \leq \ell_1,\cdots,\ell_k \leq n$. Thus g is polynomial of degree $k - 1$.

The sequence of functions $\{a_n(x)\}_{n=0}^\infty$,

$$a_0(x) = g(x)$$

$$a_1(x) = [[L_0, a_0(x)], a_0(x)]$$

$$\vdots$$

$$a_n(x) = [[L_0, a_{n-1}(x)], a_{n-1}(x)]$$

must also be in Λ. Another tedious calculation shows that

$$a_n(x) = \nabla a_{n-1}(x)\, A\, \nabla^T a_{n-1}(x) \ .$$

If $g(x)$ is a polynomial, then clearly $\{a_n(x)\}_{n=0}^{\infty}$ is a sequence of polynomials. We claim that

$$\deg a_n(x) = 2 \deg a_{n-1}(x) - 2 \ . \tag{5.48}$$

To prove this, observe $A > 0$. Hence, there exists a matrix S such that

$$SAS^T = \text{diag}\,[\lambda_1, \cdots, \lambda_n], \lambda_i > 0 \quad i = 1, \cdots n$$

so that

$$a_n(x) = \nabla a_{n-1}(x) S^{-1}\, SAS^T (S^{-1})^T \nabla^T a_{n-1}(x)$$

$$= \sum_{i=1}^{n} (\nabla a_{n-1}(x) S^{-1})_i^2 \ .$$

But $\nabla a_{n-1}(x) S^{-1}$ will be a vector of polynomials, the highest degree of which will be $(\deg a_{n-1}(x) - 1$. (5.48) clearly follows from this.

Now suppose $\deg[g(x)] > 2$. (5.48) will then imply that the sequence $\deg a_n(x)$ will increase without bound and so admit polynomials of arbitrarily high order into Λ. But this cannot happen if $\dim \Lambda < \infty$. Therefore $\deg g(x) \leq 2$.

In the scalar case, theorem 5.3 may be used to impose conditions on the drift $f(x)$ and so to obtain n.a.s.c.'s for finite-dimensionality.

Theorem 5.4 Let $n = m = p = 1$, $G = 1$. Then $\dim \Lambda < \infty$ iff

$$\text{(i)}\quad h(x) = \alpha x + \beta$$
$$f' + f^2 = ax^2 + bx + c \quad . \tag{5.49}$$

$$\text{(ii)}\quad h(x) = \alpha x^2 + \beta x + \gamma,\ \alpha \neq 0, \text{ and}$$
$$f' + f^2 = -h^2 + a(2\alpha x+\beta)^2 + b + \frac{c}{(2\alpha x+\beta)^2} \tag{5.50}$$

or

$$f' + f^2 = -h^2 + ax^2 + bx + c \quad . \tag{5.51}$$

Proof Assume $\dim \Lambda < \infty$ and $h(x) = ax + \beta$. The function

$$r_1(x) = [L_0,[L_0,h]]$$
$$= \alpha^2 x + \frac{1}{2}f'' + ff'$$

is in Λ. Hence by theorem 2,

$$r_1(x) = ex^2 + (a+\alpha^2)x - b/2$$

for some constants, e, a, and b. Thus

$$\frac{1}{2}[f'+f^2]' = \frac{1}{2} f'' + ff' = ex^2 + ax + b/2 ,$$

and hence

$$f' + f^2 = \frac{2}{3} ex^3 + ax^2 + bx + c \qquad . \qquad (5.52)$$

Likewise

$$r_2(x) = [L_0,[L_0,r_1(x)]] - 4eL_0$$

$$= 2e(2\alpha x+\beta)^2 + 2e[f'+f^2] +$$

$$(2ex+a+\alpha^2)[\tfrac{1}{2}f''+f^2+(\alpha x+\beta)]$$

is quadratic. But, by substituting (5.52) into (5.53) we find that $r_2(x)$ contains the term $e^2[\frac{7}{3}]x^3$. Hence $e = 0$. Thus $f' + f^2$ is quadratic. Conversely, if $f(x)$ satisfies (5.52) with $e = 0$,

$$A = \text{Span}\{L_0, x, \frac{d}{dx}, I\} ,$$

which is finite-dimensional.

Next suppose that $h(x) = \alpha x^2 + \beta x + \gamma$, $\alpha \neq 0$. Again

$$r(x) = [L_0,[L_0,h]] - 4\alpha L_0$$

$$= 2\alpha[f'+f^2] + \frac{1}{2}(2\alpha x+\beta)[f''+2f'f] + 2\alpha h^2$$

$$+ (2\alpha x+\beta)hh'$$

is a quadratic function $Q(x)$. Let $z = f' + f^2$. We see that z

satisfies

$$2\alpha z + \frac{1}{2}(2\alpha x+\beta)z' = - 2\alpha h^2 - (2\alpha x+\beta)hh' + Q(x) \quad . \qquad (5.54)$$

The general solution z of this equation is

$$z(x) = - h^2(x) + Q_1(x) + \frac{\lambda}{(2\alpha x+\beta)^2}$$

where $Q_1(x)$ is another quadratic function solving

$$2\alpha Q_1 + \frac{1}{2}(2\alpha x+\beta)Q_1' = Q(x)$$

and $\lambda(2\alpha x+\beta)^{-2}$ is a solution of the homogeneous part of (5.54). If $Q_1(x) \neq \mu(2\alpha x+\beta)^2 + \nu$, then by taking an appropriate linear combination of $Q(x)$ and $h(x)$ we may show that $x \in \Lambda$. It then turns out that

$$[L_0[L_0,x]] = Q_1'(x) - 2\lambda\alpha(2\alpha x+\beta)^3$$

is in Λ. But this must be quadratic and hence $\lambda = 0$.

Conversely if $Q_1(x) = \mu(2\alpha x+\beta)^2 + \nu$, then λ may be non-zero and the estimation algebra is

$$\Lambda = \text{Span}\ \{L_0,(2\alpha x+\beta)^2,(2\alpha x+\beta)\frac{d}{dx} - [\alpha-(2\alpha x+\beta)f],I\} \ .$$

If $f' + f^2 = - h^2 + ax^2 + bx + c$ and $Q_1(x)$ is not of the above special form

$$\mathcal{A} = \text{Span}\left\{L_0, x^2, x, x\frac{d}{dx} - xf, \frac{d}{dx} - f, I\right\}.$$

This completes the proof.

Remark Case (i) compares to the Beneš [1] examples; a form of (ii) appears in Brockett [6].

By a simple transformation, we may extend this result to the filtering model

$$\begin{aligned} dz(t) &= f(x(t))dt + g(z(t))db(t) \\ dy(t) &= h(z(t))dt + dw(t), \end{aligned} \tag{5.56}$$

and $h, f, g \in C^\infty(U)$, U is an open interval of $\mathbb{R}$, and $g(z) > 0$ for $z \in U$. Consider the differential equation

$$\theta'(x) = g(\theta(x)) \qquad \theta(0) = z_0 \in U.$$

Let I denote the maximal interval about $x = 0$ on which the solution $\theta(x)$ exists. θ maps I onto U, is infinitely differentiable on I, and is invertible. These statements are easily demonstrated from differential equation theory. Next let

$$\tilde{f}(x) = \frac{1}{\theta'(x)}\left[f(\theta(x)) - \frac{1}{2}\theta''(x)\right]$$

and suppose

$$dx(t) = \tilde{f}(x(t))dt + db(t) .$$

By applying Itô's rule, one finds that if $z(t) = \theta(x(t))$

$$dz(t) = f(z(t))dt + g(z(t))db(t) .$$

Hence, (5.56) may be replaced by:

$$dx(t) = \tilde{f}(x(t))dt + db(t) \qquad (5.57)$$

$$dy(t) = h(\alpha(x(t)))dt + dw(t).$$

The Lie algebra analysis can now be carried out on (5.57). The estimation algebra is $\Lambda_1 = \{\frac{1}{2}\frac{\partial^2}{\partial z^2} g^2(t) - \frac{\partial}{\partial t} f(z) - \frac{1}{2} h^2(z), h(z)\}_{LA}$, that of (5.57) is $\Lambda_2 = \{\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{\partial}{\partial x}\tilde{f}(x) - \frac{1}{2}(h\circ\alpha)^2, h\circ\alpha\}_{LA}$.

If $i:\Lambda_2 \to \Lambda_1$ is defined by

$$(i[B]\psi)(z) = B\psi\circ\theta(x)\Big|_{x=\theta^{-1}(z)}$$

one sees easily that i is an isomorphism of Λ_1 and Λ_2. Hence, we derive.

Theorem 5.5 $\quad \dim \Lambda_1 < \infty$ iff $\dim \Lambda_2 < \infty$.

Theorem 5.5 says that any finite dimensional estimation

algebra for the model (5.56) with $g(z) > 0$ can be reduced by state space diffeomorphism to one of the cases in theorem 5.4.

Solution of Zakai's Equation

Which cases among those singled out in theorem 2, allow a solution of the filtering problem via the method of section 5. ? To answer this, it is first important to characterize the drifts f solving the equations (5.49) - (5.51). It turns out that f may explode for finite x, that is, the maximal interval $U = (r_0, r_1)$ on which f can be defined may be only bounded or semi-infinite because $|f(x)| \to \infty$ as $x \to$ bounded endpoint of U. Despite this a signal solving

$$dx(t) = f(x(t))\, dt + db(t)$$

$$x(0) \in U \tag{5.58}$$

will still exist, but, in general only up to a stopping time, τ, at which it attains a boundary point of U, (Gihman and Skorohod [17]). The theory of diffusions on bounded intervals must now be applied to proceed further. It says that to specify $x(t)$ for $t > \tau$, one must impose conditions that tell how the process behaves at the boundary, i.e., whether it is absorbed, reflected, terminated, or some combination of these three. Always, when $x(t) \in U$, it is assumed to solve (5.58). The theory also

indicates that even if $x(t)$ does not hit the boundary, so that $x(t)$ can be defined as a solution of (5.58) without additional conditions, different process behavior near the boundaries can occur. Whether or not $x(t)$ attains the boundary and how it acts near a boundary depends on the nature of f.

To study f and (5.58) we state some preliminary results from diffusion theory. In our definitions and statements we follow the exposition of Mandl [27]; the original reference is Feller [12]. Let $f \in C^1(U)$, $U=(r_0,r_1)$ and consider the operator

$$B = \frac{1}{2} \frac{d^2}{dx^2} + f(x) \frac{d}{dx} .$$

Let $r \in (r_0,r_1)$ and define

$$c(x) = 2 \int_r^x f(s)ds$$

$$u(x) = \int_r^x dy\, e^{-c(y)} \int_r^y dz\, e^{c(z)}$$

$$v(x) = \int_r^x dy\, e^{c(y)} \int_r^y dz\, e^{-c(z)}$$

Definition 5.3 For the operator B,

The boundary point r_i is

(a) an inaccessible, natural boundary if $u(r_i)=\infty$, $v(r_i)=\infty$

(b) an inaccessible, entrance boundary if $u(r_i)=\infty$, $v(r_i)<\infty$

(c) an accessible, exit boundary if $u(r_i)<\infty$, $v(r_i)<\infty$.

(d) an accessible, regular boundary if $u(r_i), v(r_i) < \infty$.

Let $C(\overline{U}) := \{\psi(x), x\in U \mid \psi$ is continuous, $\lim_{x\to r_i} \psi(x)$ exists and is finite, $i = 0,1\}$

and consider the differential operation B

on

$$D(B) = \{\psi \in C(\overline{U}) \mid B\psi \in C(\overline{U})\}$$

Lemma 5.3

i) If the boundaries of U are inaccessible then B on D(B) generates a unique Markov semigroup on $C(\overline{U})$

ii) Suppose B has an accessible boundary. Define D'(B) as the set of $\psi \in D(B)$ s.t. at a regular boundary r_i

$$0 = p_i\psi(r_i) - (-1)^i \tau_i \lim_{x\to r_i} e^{(x)} \psi(x) + \sigma_i B\psi(r_i)$$

$$\sigma_i + \tau_i > 0,\ p_i \tau_i,\ \sigma_i \geq 0$$

and at an exit boundary $0 = p_i v(r_i) + \sigma_i Bv(r_i)$, $p_i \sigma_i \geq 0$, $p_i + \sigma_i > 0$. Then B on $D'(B)$ will generate a Markov semigroup on $C(\overline{U})$.

Proof. Mandl [27].

Remark. The boundary conditions in lemma 5.5 ii) are called local. More general, "lateral" conditions, corresponding to having the process jump to a point inside U, are possible, but for such signals Zakai's equation no longer holds.

For a domain with inaccessible boundaries, let $x(t;x_0)$ denote the solution of (5.58) with $x(0;x_0) = x_0 \in U$. Then B on $D(B)$ generates $(T(t)g)(x_0) = Eg(x(t;x_0))$ for $g \in C(\overline{U})$. Construction of Markov processes that correspond to the semigroup generated in the case of accessible boundaries with local boundary conditions may be found in Gihman and Skorohod [17].

We shall now analyze the solutions f of (5.49)-(5.51) in terms of this theory. Given a function f, let $v = \exp \int^x f$, or, vice-versa, given v, let $f = \dot{v}/v$. Then f satisfies (5.49), (5.50) or (5.51) respectively, if and only if v satisfies

$$v'' = (ax^2 + bx + c)\, v \tag{5.59}$$

or

$$v'' = \left(-\tfrac{1}{4}(2ax+b)^2 + a(2ax+b)^{\frac{1}{2}} - \frac{3}{2(2ax+b)^{\frac{1}{2}}}\right) v. \tag{5.60}$$

or

$$v'' = (-h^2 + ax^2 + bx + c)\, v \tag{5.61}$$

respectively. Clearly, singularities of f will arise at zeroes of v. The situation is summarized in the following lemma. Note that special attention must be given to (5.60) because of the singularity of the coefficient at $x_0 = -b/2a$.

Lemma 5.4

i) Suppose that f is a solution of (5.49) or (5.51) and that f becomes singular at x_0. Then $f(x) = \frac{1}{x-x_0} + O(1)$ as $x \to x_0$. The same holds true if f solves (5.50) and $x_0 \neq \frac{-b}{2a}$.

ii) If f is a solution (5.50) and $c \geq -\frac{1}{4}$, then f can have a solution on an interval with endpoint $x_0 = \frac{-b}{2a}$ and

$$f(x) = \frac{\alpha}{x-x_0} + O(1) \quad , \quad x \to x_0$$

where $\alpha(\alpha-1)=c$.

iii) If $c < -\frac{1}{4}$, f is not defined on an interval with endpoint x_0 or containing x_0.

Proof (i) Let $v(x_0) = 0$. Then $v'(x_0) \neq 0$ (or otherwise $v \equiv 0$).

Thus

$$f(x) = \frac{v'(x_0)+v''(x_0)(x-x_0)+\cdots}{v'(x_0)(x-x_0)+\frac{1}{2}v''(x_0)(x-x_0)^2+\cdots}$$

$$= \frac{1}{x-x_0} + O(1) \quad , \; x \to x_0 \; .$$

(ii) It suffices to consider the case $a = 0$ and $\alpha = 1$, for which $x_0 = 0$, and

$$v'' = (-h^2 + ax^2 + b + \frac{c}{x^2})v \quad .$$

If $c > -\frac{1}{4}$ $\alpha(\alpha-1) = c$ has two solutions $\alpha_1 < \frac{1}{2} < \alpha_2$ and v has series solutions near zero

$$v_1(x) = x^{\alpha_1} \sum_{n=0}^{\infty} a_n x^n, \quad a_0 \neq 0$$

$$v_2(x) = x^{\alpha_2} \sum_{n=0}^{\infty} a_n x^n, \quad b_0 \neq 0 \quad .$$

For each solution

$$f = \frac{\alpha_i}{x_i} + O(1) \quad , \; x \to 0$$

For $c = -1/4$, the solutions v are generally of the form

$$v_1(x) = x^{1/2} \sum_{n=0}^{\infty} a_n x^n$$

$$v_2(x) = (\ln x)\, x^{1/2} \left(\sum_{n=0}^{\infty} a_n x^n \right) + x^{1/2} \sum_{n=0}^{\infty} b_n x^n \quad .$$

Again

$$f = \frac{1}{2}\frac{1}{x} + O(1) \quad .$$

(iii) If $c < -\frac{1}{4}$, the solutions v_1 and v_2 are still valid, but now α_1 and α_2 are complex

$$\alpha_i = \frac{1}{2} \pm i\, \frac{\sqrt{|4c+1|}}{2} \quad .$$

Thus since

$$x^{1/2 \pm i \frac{1}{2}\sqrt{|4c+1|}} = x^{1/2}\, e^{\pm i \frac{1}{2}\sqrt{4c+1}\ln x}$$

any real solution v will have an infinite number of zeroes in any neighborhood of zero. Since each of these zeroes will correspond to a singularity of f, f cannot be well-defined in an interval containing 0 or with 0 as an endpoint.

From lemma 5.4 we can derive

Lemma 5.5

Let f be a solution of any of the equations (5.49) - (5.51) on $U = (r_0, r_1)$. If r_i is finite and $\lim_{x \to r_i} |f(r_i)| < \infty$, r_i is regular. If r_i is finite and $f(x) = \frac{\alpha}{x - r_i} + {}^{x \to r_i}O(1)$ $(x \to r_i)$ r_i is an entrance boundary if $\alpha \geq 1/2$, a regular boundary if $\alpha \in (-1/2, 1/2)$, and an exit boundary of $\alpha \leq -1/2$. If r_i is infinite ($+\infty$ or $-\infty$), it is natural.

Proof For the finite boundaries, calculate $u(x)$ and $v(x)$ and apply definition 5.3. To prove that an infinite boundary is natural it is necessary to know how f behaves at that boundary. It will be shown later that f can exist in a semi-infinite or infinite domain only if it solves (5.49). Suppose for instance that $r_1 = +\infty$. It turns out that

$$f' + f^2 = ax^2 + bx + c$$

whether either $a > 0$ or $a = 0$, $b > 0$, or $a = 0$, $b = 0$, $c \geq 0$. In any of these cases $f(x)$ can grow at most like x as $x \to \infty$ and from this one can show $\lim_{x \to \infty} u(x) = +\infty$ $\lim_{x \to \infty} v(x) = +\infty$.

Let us take up the question raised in the last proof; which equations of (5.49)-(5.51) have solutions on which kinds of domains?

Lemma 5.5

Suppose f satisfies (5.50), (5.51) or $f'+f^2=ax^2+bx+c$ and either $a<0$, or $a=b=0$, $c<0$. If f is defined on $U=(r_0,r_1)$ and has no singularities in U, then U must be bounded. If f satisfies $f'+f^2=ax^2+bx+c$ and r_i is infinite, $\lim_{x\to r_i} ax^2+bx+c \geq 0$.

Proof. The proof applies the Sturm-Liouville comparison theorem (see, e.g. Coddington and Levinson [8]). Suppose f satisfies (5.51). Then $v = \exp \int_r^x F(x)$ satisfies

$$v'' + (h^2-ax^2-bx-c)\, v=0$$

Since h is quadratic, for $|x|$ large enough

$$h^2(x) - ax^2 - bx^2 - c > 1.$$

Hence, by the Sturm-Liouville comparison theorm v must have a zero between any two successive zeroes of any solution of

$$\vartheta'' + \vartheta = 0.$$

Since $\vartheta = c_1\cos x + c_2\sin x$ we see that v must have a zero in any unbounded domain. Thus f can be defined without singularities only in unbounded domains. The other statements are proved analogously.

The interesting results contained in these lemmas are (1) that no finite boundary arising from an f solving (5.49), (5.50),

or (5.51) will be natural and (2) that finite boundaries will always be encountered, except possibly, with (5.49). It will turn out that these non-natural boundaries will generate non-trivial boundary conditions for Zakai's equation. The general implication is then that, just as in the absorbed Brownian motion example, these boundary conditions will interfere with the attempt to apply Lie algebraic techniques. It would then follow that for scalar-signal models of the type (5.58) only the known examples of Beneš can be treated by the method of section 5.1.

It is difficult to prove this statement in such sweeping generality. First, there is a problem of formulation. If $x(t)$ solves:

$$dx(t) = f(x(t))\,dt + db(t)$$

on a finite domain U with regular accessible boundaries, it is no longer possible to characterize the conditional distribution just by a density $p(x,t)$ on U; one must also consider the conditional mass distributions $P_{r_i}(t) = E_0\{1_{(x(t)=r_i)} L_t | \mathcal{F}_t^y\}$. For general local boundary conditions, we must solve a system of equations for the triple $(P_{r_0}(t), p(x,t), P_{r_1}(t))$, and Lie algebra techniques, if any, must be applied to this system. Thus in the analysis of Lie algebraic techniques to follow we will avoid the accessible boundary case.

Remark. For certain boundary conditions in the regular accessible case the system $(P_{r_0}, p(x,t), P_{r_1}(t))$ degenerates into just a density equation with boundary condition; if r_i is pure reflecting, i.e., $\phi \in D'(B)$ only if $\frac{\partial}{\partial x}\phi|_{x=r_i} = 0$, then $P_{r_i}(t) \equiv 0$ and $(\frac{\partial}{\partial x} - f(x))p(x,t)|_{x=r_i} = 0$, $\forall t$; if r_i is pure absorbing, i.e., $\phi \in D'(B)$ iff $B\phi(r_i) = 0$, then $dP_{r_i}(t) = [\frac{\partial}{\partial x} p(x,t)|_{x=r_i}]dt$ and $p(r_i,t) = 0$, $\forall t$. (see Pardoux [40]). Theorem 5.5 will hold for these cases, but we will not work out the details.

Secondly, the case

$$f' + f^2 = -h^2 + a(2\alpha x+\beta)^2 + b + \frac{c}{(2\alpha x+\beta)^2} \quad \begin{matrix} c \neq 0 \\ c \geq -4 \end{matrix}$$

on an interval with $x_0 = -\beta/2\alpha$ as an endpoint poses analytical difficulties, both because of the variety of cases $f(x) \sim \frac{\alpha}{x-x_0}$, $x \to x_0$, $\alpha \in \mathbb{R}$, to be analyzed and because of the singularity $\frac{c}{2\alpha x+\beta^2}$ in $f' + f^2$. In the interests of simplicity, we will not treat this case.

From now on, we will be interested in solving Zakai's equation for the problem of filtering a signal $x(t)$ solving (5.58) under the assumptions:

(i) h is linear or quadratic

(ii) f solves one of (5.49) - (5.51)

(iii) $U = (r_0, r_1)$ is the full domain of definition of f (H)

iv) If f solves (5.50) with $c \neq 0$, $c \geq -\frac{1}{I}$, neither endpoint r_i equals $x_0 = -\beta/2\alpha$

From the lemmas presented above, one can conclude that if r_i is a finite boundary, it is of entrance type, and $f(x) \sim \frac{1}{x-r_i}$ as $x \to r_i$.

Zakai's equation for any one of these problems is:

$$\frac{\partial}{\partial t} p(x,t) = \{\frac{1}{2} \frac{\partial^2}{\partial x^2} - \frac{\partial}{\partial x} f(x) - \frac{1}{2} h^2(x)\} p(x,t) + \dot{y}(t)h(x)p(x,t) \tag{5.62}$$

$$\lim_{x \to r_i} \{\frac{\partial}{\partial x} - 2f(x)\} p(x,t) = 0 \quad \text{if } r_i \text{ is a finite (entrance) boundary.} \tag{5.63}$$

We give a brief, formal derivation of this for the case $r_0, r_1 < \infty$.

For $g \in C(\bar{U})$ define

$$\sigma_t(g) = E_0\{g(x_t)L_t|F_t^y\} = \int_{r_0}^{r_1} g(z)p(z,t)dz \ .$$

If $g \in D(B)$,

$$d\sigma_t(g) = \sigma_t(Bg)dt + \sigma_t(hg)dy(t) \quad , \tag{5.64}$$

(Kunita [25]). By integrating the term $\sigma_t(Bg)$ by parts we derive

$$d\sigma_t(g) = \int_{r_0}^{r_1} g(z)dp(z,t)\,dz$$

$$= \int_{r_0}^{r_1} g(z)[B^*p(z,t)dt + h(z)p(z,t)dy(t)]dz$$

$$+ \frac{1}{2} e^{c(x)}g'(x)e^{-c(x)}p(x,t)\Big|_{r_0+}^{r_1-} -$$

$$g(x)[\frac{\partial}{\partial x} - 2f(x)]p(x,t)\Big|_{r_0+}^{r_1-}$$

$$B^* = \frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{\partial}{\partial x} f(x) \ .$$

In order that this hold for general $g \in D(B)$, we require

$$dp(x,t) = B^*p(x,t)dt - h(x)p(x,t)dy(t) \tag{5.65}$$

$$\lim_{x \to r_i} g'(x)p(x,t) = 0 \tag{5.66}$$

$$\lim_{x \to r_i} g(x)[\frac{\partial}{\partial x} - 2f(x)]p(x,t) = 0 \quad . \tag{5.67}$$

For $g \in D(B)$ one can easily show that

$$\lim_{x \to r_i} \{\frac{\partial}{\partial x} - 2f(x)\}p(x,t) = 0$$

implies both (5.67) and (5.66). When writing (5.65) with $\dot{y}(t)$ instead of $dy(t)$ one must add the Wong-Zakai correction term $-1/2\ h^2(x)$ to B^*. This proves (5.62) - (5.63).

Rather than work with (5.62) directly it is convenient to invoke the "gauge" transformation

$$e^{F(x)} = e^{\int_r^x f(z)dz} \qquad r \in (r_0, r_1)$$

$$\left(= e^{1/2\alpha(x)} \right)$$

and write

$$p(x,t) = e^{-F(x)}q(x,t) \quad .$$

A calculation shows that

$$\frac{\partial q}{\partial t} = \{\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}[f' - f^2 + h^2]\}q(x,t) + \dot{y}(t)h(x)q(x,t) \tag{5.68}$$

$$\lim_{x \to r_i} e^{F(x)} \left[\frac{\partial}{\partial x} - f(x)\right] q(x,t) = 0 \quad \text{if } r_i \text{ is an entrance boundary} \tag{5.69}$$

It is this equation that we seek to solve by Lie algebraic methods. Let $A = \frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}[f' + f^2 + h^2]$ and call $\{A,h\}_{LA}$ the algebra of (5.68). $\{A,h\}_{LA}$ is isomorphic to $\{L_0,h\}_{LA}$ by the map $B \to e^{F} B e^{-F}$ from $\{A,h\}_{LA}$ to $\{L_0,h\}_{LA}$. This map also translates evolutions; that is if

$$\lim_{t \downarrow 0} \left[\frac{e^{tA} - I}{t}\right]\phi = A\phi \qquad \phi \in D$$

then

$$\lim_{t \downarrow 0} \frac{1}{t}\left[e^{F} e^{tA} e^{-F} - I\right]\phi = L_0\phi, \phi \in e^{F} D \quad .$$

Thus the Lie algebra analysis may be carried out on (5.68).

The technique we want to explore is that of solving (5.68) by

$$q(x,t) = (e^{g_1(t)A} e^{g_2(t)X_2} \cdots e^{g_d(t)X_d} q_0)(x) \; . \tag{5.70}$$

We place $e^{g_1(t)A}$ first in this series because we need an evolution that must, in general, satisfy boundary conditions. It is not usually possible to do this with first order operators, and A is of second order. Different second order operators in $\{A,h\}_{LA}$ might be chosen instead of A, but this will make no difference in what follows.

The natural semigroup to associate with A is that determined by the boundary conditions (5.69). We will define this semigroup on $L^2(U)$. Thus let

$$D(A) = \{\psi \epsilon L^2(U) \mid A\psi \epsilon L^2(U),\ \psi(r_i) = 0,\ r_i \text{ an entrance boundary}\} \quad .$$

Assume r_i is an entrance boundary. The condition $\psi(r_i) = 0$ does not look like (5.69), but it is in fact equivalent. Since $A\psi = \varphi \in L^2(U)$,

$$\frac{d}{dx}\psi(x) = \frac{d\psi}{dx}(x)\Big|_{x=r} + \int_r^x [f' + f^2 + h^2](s)\psi(s)ds \quad .$$

Since $e^{F(x)} \sim kx$ $(x \to r_i)$ for some constant k we see that

$$\lim_{x \to r_i} e^{F(x)} \frac{d}{dx}\psi(x) = 0 \quad .$$

Thus $\lim_{x \to r_i} e^{F(x)} \left[\frac{d}{dx} - f(x)\right]\psi(x) = \lim_{x \to r_i} e^{F(x)} f(x)\ \psi(x) =$

$$= \lim_{x \to r_i} \frac{kx}{x}\psi(x) = k\psi(r_i) = 0.$$

We now state the main theorem.

Theorem 5.5

Let η, f and U satisfy the hypotheses (4), and suppose

U is bounded or semi-infinite

(i) $-A$ is self-adjoint and bounded-below on $D(A)$ and hence generates a semigroup e^{tA}. Also, A has a discrete spectrum.

(ii) Let $D'(A) = \{$finite linear combinations of eigenvalues of $A\}$

$$e^{tA} h\, \phi \neq \Big(\sum_{n=0}^{\infty} \frac{t^n}{n!} [\text{ad } A]^n\, h \Big) e^{tA} \phi \quad \text{for } \phi \in D'(A),\ \phi \neq 0 . \tag{5.71}$$

Proof Assume, without loss of generality, that $h(x) = x$ in the linear case, and $h(x) = x^2$ in the quadratic case. It is easiest to prove the theorem separately for the cases U bounded and U semi-infinite.

Let U be bounded. Then

$$A = \frac{1}{2} \frac{d^2}{dx^2} - \frac{1}{2} [f' + f^2 + h^2]$$

and $\frac{1}{2}[f'+f^2+h^2]$ is a bounded, C^∞ function on U. (Recall that (H) excludes the case in which $f' + f^2$ may have a singularity in $\overline{U}$.) A standard calculation from differential equation theory (Coddington and Levinson [3]) shows that A is self-adjoint on $D(A)$ and $-A$ is bounded below. Further, the theory of self-adjoint, differential operators on bounded intervals implies that A on $D(A)$ has a discrete spectrum λ_n, s.t., $\lambda_n > \lambda_{n-1}$, $\lim_{n\to\infty} \lambda_n = -\infty$, and $\sum_{n=1}^{\infty} \frac{1}{|\lambda_n|^2} < \infty$, and a corresponding complete set of orthogonal eigen-

vectors $\{\psi_n\}_{n=0}^{\infty}$. (Coddington and Levinson [8]). If $\phi \in L^2(\bar{U})$

$$(e^{tA}\phi)(x) = \int_U [\sum_{n=0}^{\infty} e^{-\lambda_n t}\psi_n(x)\psi_n(y)]\phi(y)dy \qquad (5.72)$$

(Coddington and Levinson [8]). One now proves (ii) case by case. To illustrate we do the example $h = x^2$ and $f' + f^2 = -h^2 + ax^2 + c$. The estimation algebra $\{A,h\}_{LA}$ is then spanned by $\{A,x^2,x\frac{d}{dx}, I\}$ and has commutation relations

$$[A, x^2] = 2x\frac{d}{dx} + 1$$

$$[A, x\frac{d}{dx}] = 2A + 2a\,x^2 + c \quad .$$

From these, one readily derives

$$\sum_{n=0}^{\infty} \frac{t^n}{n!}[\text{ad } A]^n h = \frac{1}{a}[\cosh 2\sqrt{a}t - 1]A + [\cosh 2\sqrt{a}t]x^2$$

$$+ \frac{1}{\sqrt{a}}(\sinh 2\sqrt{a}t)x\frac{d}{dx} + g(t)$$

$$g(t) = \frac{1}{2\sqrt{a}}\sinh 2\sqrt{a}t + \frac{c}{2a}\cosh 2\sqrt{a}t - \frac{c}{2a} \quad .$$

Now let us check (5.72) on ψ_n. Observe first that

$$e^{tA}\psi_n = e^{\lambda_n t}\psi_n$$

and $\psi_n'(r_i) \neq 0$. (Otherwise $\psi_n(r_i) = 0$, $\psi_n''(r_i) = 0$ would imply $\psi_n \equiv 0$). Thus

$$\left(\sum_{n=0}^{\infty} \frac{t^n}{n!} [\text{ad } A]^n h\right) e^{tA} \psi_n(x)$$

$$= e^{\lambda_n t} a^{-1/2} (\sinh 2\sqrt{a}\, t)x\, \psi_n'(x) \tag{5.73}$$

$$+ (\lambda_n a^{-1}(\cos 2\sqrt{a}\, t - 1) + g(t) + x^2 \cosh 2\sqrt{a}\, t) e^{\lambda_n t} \psi_n(x) .$$

This will not satisfy the boundary conditions $\psi(r_i) = 0, i = 0,1$ because of the $\psi_n'(x)$ term. Thus (5.73) cannot equal $e^{tA} h\psi$ since by (5.72), $(e^{tA} h\psi)(r_i) = 0, i = 0,1$. Now let $\psi(x) = \sum_{n=1}^{N} \alpha_n \psi_n$, $N < \infty$. Then

$$\left(\sum_{n=0}^{\infty} \frac{t^n}{n!} [\text{ad } A]^n h\right) e^{tA} \psi(r_i) = 0, \quad i = 0,1$$

only if

$$\sum_{j=0}^{N} \alpha_j e^{\lambda_j t} = 0 ,$$

(see 5.73). Since the λ_j are unequal this cannot happen unless $\alpha_j = 0$, $j = 1, \cdots, n$. The statement analogous to (ii) with n replaced by $2x \frac{d}{dx}$ is also true. The other cases involving bounded U are proved in the same manner.

Now let $U = (r_0, \infty)$, r_0 finite. As was shown above this

case can only occur if $f' + f^2 = ax^2 + bx + c$ and $a \geq 0$ at the very least. Hence the operator A is of the form

$$\frac{1}{2}\frac{d^2}{dx^2} - \frac{1}{2}[x^2 + ax^2 + bx + c]$$

and $\lim_{x\to\infty} \frac{1}{2}[x^2+ax^2+bx+c] = \infty$. It is then known that A is self-adjoint on $D(A) = \{\psi | A\psi \in L^2(U), \psi(r_0) = 0\}$ with a strictly decreasing sequence of eigenvalues λ_j. The analysis then proceeds exactly as for the absorbed Brownian motion case. (Indeed by a transformation $z = \nu x + s$, $\dot{y}'(t) = \theta\dot{y}(t)$, (5.68) with boundary condition $q(0,t) = 0$ becomes the Zakai equation for absorbed Brownian motion). The proof for the case $U = (-\infty, r_1)$ $r_1 < \infty$, is the same.

Remarks

(i) Theorem 5.9 implies that the crucial lie algebra identity (I) of section 5.1 fails, and hence that the method (5.70) of solving (5.68) will not work. The result is that, exclusive of the cases not satisfying the assumptions (H) the Lie algebra technique works only for the case treated by Beneš in which f is a global solution of

$$f' + f^2 = ax^2 + bx + c.$$

(ii) The one non-rigorous point in the above calculation was the derivation of Zakai's equation. However, the equation (5.68) for $q(x,t)$ can be rigorously derived, if the differentiability of $q(x,t$

is assumed. Indeed if $(\Omega,\mathcal{F},\tilde{P})$ is a probability space on which $B(t)$ and $y(t)$ are independent Brownian motions; it can be shown by change of measure techniques that

$$q(x,t) = \tilde{E}\{\exp[\int_0^t h(x_0+b(s))dy(s) - \frac{1}{2}\int_0^t [f'+f^2+h^2](x_0+b(s))ds]$$

$$\times 1_{\{t<\tau\}}|\mathcal{F}_t^y, B(t)+x_0=x\}r(x,t)$$

where $\tau = \inf\{t|x_0+B(s) \in U^c\}$ and $r(x,t)$ is the density of $B(t) + x_0$ (Liptser and Shiryayev [28]). The techniques of Zakai [47] applied to this object yield (5.68),and the boundary conditions $q(r_i,t) = 0$ at finite boundaries r_i are clear. The details will not be presented.

It is in trying to justify Zakai's equation rigorously that the singularity $\frac{c}{(2ax+g)^2}$ at $x_0 = g/2a$ causes trouble and is why we excluded the case from the analysis of theorem 5.5. Nevertheless, if one presumes Zakai's equation holds and applies the method of theorem 5.5 with careful attention to what happens at x_0, the same conclusion about $e^{tA}h$ will follow.

(iii) From the remark on page 158 , the analysis of theorem 5.5 can be applied to the case when any one of the boundaries is regular and either instantaneously reflecting, on purely absorbing.

(iv) In theorem 5.5, as in the absorbed Brownian motion case of section 2, if $\mathcal{D} \subset \mathcal{D}(A)$ is invariant under A and η, it will not contain an invariant set of analytic vectors for A and η.

In our analysis of Lie algebraic techniques we excluded a number of possible, finite dimensional examples by restricting ourselves to problems satisfying hypothesis (H). However, remarks (i) - (iv) state that the results of theorem 5.5 are true, at least by formal arguments, for any excluded case for which the conditional distribution can be found by solving Zakai's equation for with (possible) boundary conditions. Thus, we conjecture that no other examples beyond those of Beneš can, in fact, be solved by the method developed in this chapter.

BIBLIOGRAPHY

1. Benes, V., "Exact Finite Dimensional Filters for Certain Diffusions with Nonlinear Drift," to appear in Stochastics, 1980.

2. Brockett, R. W., "Lie Algebras and Lie Groups in Control Theory," In Geometric Methods in Control Theory, (D.O. Mayne and R. W. Brockett, e.d) Reidel, Dordrecht, Holland (1973).

3. Brockett, R. W., "Volterra Series and Geometric Control Theory," Automatica, Vol. 12, 1976, pp. 167-176.

4. Brockett, R. W. "Nonlinear Systems and Differential Geometry," Proceedings of the IEEE, Jan. 1976, pp. 61-72.

5. Brockett, R. W., "Some Remarks on Finite Dimensional Nonlinear Estimation," Journees sur L'Analyse des Systems, Bordeaux, 1978.

6. Brockett, R. W. "Classification and Equivalence in Control Theory," Proceedings, 18th IEEE Conference on Decision and Control, 1979, Ft. Lauderdale, Fla.

7. Brockett, R. W., and Clark, J. M. C., "Geometry of the Conditional Density Equations," Proceedings, Int. Conf. on Analysis and Optimization of Stochastic Systems.

8. Coddington, E. and Levinson, N., Theory of Ordinary Differential Equations, McGraw Hill, New York, 1955.

9. Davis, M. H. A., Linear Estimation and Stochastic Control, Chapman and Hall, London, 1977.

10. Davis, M. H. A., "A Direct Proof of Innovations/Observations Equivalence for Gaussian Processes," IEEE Trans. Information Theory, Vol. 24, 2, March, 1978.

11. Eterno, J. S., Nonlinear Estimation and Phase-Lock Loops, PhD Dissertation, Dept. of Aeronautics and Astronautics, 1976, MIT.

12. Feller, W., "The Parabolic Differential Equations and the Associated Semigroup of Transformations," Ann. Math., Vol. 55, 1952, pp. 468-519.

13. Flato, Simon, Dellman and Sternheimer, "Simple Facts About Analytic Vectors and Integrability," Ann. Scient. Ec. Norm. Sup. 4e série, t5, 1972, 423-434.

14. Fleming, W. and Rishel, R., Deterministic and Stochastic Optimal Control, Springer-Verlag, New York, 1975.

15. Fujisaki, Kallianpur and Kunita, "Stochastic Differential Equations for the Non Linear Filtering Problem," Osaka J. Math., 9, 1972, 19-40.

16. Geesey, R., Canonical Representations of Second Order Processes with Applications, PhD Dissertation, Department of Electrical Engineering, Stanford University, 1969.

17. Gihman, I. I. and Skorohod, A. V., Stochastic Differential Equations, Springer-Verlag, New York, 1972.

18. Hida, T. "Nonlinear Brownian Functionals", Proceedings 18th IEEE Conference on Decision and Control, Ft. Laucerdale, Fla. 1979.

19. Hida, T. and Kallianpur, G., "The Square of a Gaussian Markov Process and Nonlinear Prediction," J. of Mult. Anal. 5, 1977, pp451-461.

20. Ito, K., "Multiple Wiener Integrals," J. Math. Soc., Japan, 3, 1951, pp. 157-169.

21. Kallianpur, G., "The Role of Reproducing Kernel Hilbert Spaces in the Study of Gaussian Processes," in Advances in Probability and Related Topics, P. Neg, ed., New York, Marcel Dekker, 1970, 49-83.

22. Kallianpur, G., and Striebel, C., "Estimation of Stochastic Systems: Arbitrary System Process with Additive White Noise Observation Errors," Ann. Math. Stat., 39, 1968, pp. 785-801.

23. Kallianpur, G. and Striebel C., "Stochastic Differential Equations in Stochastic Estimation Problems," Multivariate Analysis II, Academic Press, 1969.

24. Kalman, R. and Bucy, R., "New Results in Linear Filtering and Prediction Theory," J. Basic Engr. (Trans. ASME), Vol. 83-D, 1961, 95-108.

25. Kunita, H., "Asymptotic Behavior of the Nonlinear Filtering Errors of Markov Processes, Journal of Multivariate Analysis, 1, 1971, pp. 1-11.

26. Magnus, W., and Oberhettinger, F., Formulas and Theorems for the Special Functions of Mathematical Physics, Springer Verlag, New York 1966.

27. Mandl, P., Analytical Treatment of One-Dimensional Markov Processes, Springer Verlag, 1968.

28. Liptser and Shiryayev, Statistics of Random Processes, I and II, Springer Verlag, 1977.

29 Marcus, S. T. Estimation and Analysis of Nonlinear Stochastic Systems, Report ESL-R-601, Electronic Systems Laboratory, MIT, 1975.

30. Marcus, S., Mitter, S. K. and Ocone, D., "Finite Dimensional Nonlinear Estimation in Continuous and Discrete Time," Proceedings, Int. Conf. on Anal. and Opt. of Stochastic Systems, Oxford, 1978.

31. Marcus and Willsky, "Algebraic Structure and Finite Dimensional Nonlinear Estimation", SIAM. J. Math. Anal. 9 (2), 1978, pp. 312-325.

32. McKean, H. P., "Wiener's Theory of Nonlinear Noise," Stochastic Differential Equations, SIAM-AMS Proceedings, American Math. Soc., Providence, R. I., 1973, pp. 191-209.

33. Meyer, P. A., Un cours sur les integrales stochastiques in Seminaire de Probabilites X, Springer Lecture Notes 511, 1976.

34. Miller, K. S. Multidimensional Gaussian Distributions, John Wiley, 1965.

35. Mitter, S. K., "Filtering Theory and Quantum Fields," to appear in Asterisque.

36. Mitter, S. K. and Ocone, D., "Multiple Integral Expansions for Nonlinear Filtering, " Proceedings of the 18th IEEE Conference on Decision and Control, Ft. Lauderdale, Fla. 1979.

37. Nelson, E., "Analytic Vectors," Ann. Math., 70, 1959, pp. 572-615.

38. Nelson, E., "The Free Markov Field", J. Functional Analysis, 12, 1973, pp. 211-227.

39. Neveu, J., "Sur l'esperance conditionelle par rapport a un mouvement brownien," Ann. Inst. Henri Poincare, Section B, Vol. XII, 2, 1976, pp. 105-109.

40. Pardoux, E., "Stochastic Partial Differential Equations and Filtering of Diffusion Processes," Stochastics, Vol. 3, 1979, pp. 127-167.

41. Reed, M. and Simon, B., Methods of Mathematical Physics I: Functional Analysis, Academic Press, New York, 1972.

42. Reed, M. and Simon, B., Methods of Mathematical Physics II: Fourier Analysis, Self-Adjointness, Academic Press, New York, 1975.

43. Segal, I., "Construction of Non-Linear Local Quantum Processes: I", Ann. Math., Vol. 92, No. 3, 1970, 462-481.

44. Wei, J. and Norman E., "On The Global Representation of the Solutions of Linear Differential Equations as a Product of Exponentials," Proc. Am. Math. Soc., April, 1964.

45. Wiener, N., Nonlinear Problems in Random Theory, Wiley, New York, 1958.

46. Wong, E., Stochastic Processes in Information and Dynamical Systems, McGraw-Hill, 1972.

47. Zakai, M., "On the Optimal Filtering of Diffusion Processes," Z. Wahr, verw. Geb., 11, 1969, pp. 230-243.

48. Van Trees, H. L., Detection, Estimation, and Modulation Theory, New York, Wiley, 1968-1971, 3v.

49. Lamperti, Stochastic Processes, Applied Mathematical Sciences, Vol. 23, Springer-Verlag, New York, 1977.

APPENDIX 1 HOMOGENEOUS CHAOS THEORY

Wiener's homogeneous chaos theory provides a method of representing functions of Gaussian processes by certain infinite expansions, the terms of which are the prototypes of multiple stochastic integrals. This appendix presents the fundamental motions of the theory; the treatment follows that of Kallianpur [21].

The situation is as follows. Let T be a separable, topological space. $\{x(t)|t \in T\}$ will denote a mean zero Gaussian process on T with a covariance function $c(t,s) = Ex(t)x(s)$ that is jointly continuous in t and s. Let the probability space of the process be (Ω, F, P) and let $A = \sigma\{x(s)|s \in T\}$. Homogeneous chaos theory concerns itself with the structure of $L^2(\Omega,A,P)$ considered as a Hilbert space with inner product $\langle \phi,\psi \rangle = E[\phi\psi]$. It seeks to build a useful orthonormal basis for $L^2(\Omega,A,P)$.

We present the basic construction. Let

$$H = \overline{\text{Span}} \left\{ \sum_1^N \alpha_i x(t_i) | t_i \in T, N < \infty \right\}$$

('——' denotes closure in the norm of $L^2(\Omega,A,P)$). H is the subspace of $L^2(\Omega,A,P)$ consisting of linear functionals of $x(\cdot)$. H is separable, because of the separability of T and the continunity of $c(t,s)$, and hence H has a countable orthonominal basis $\{\xi_i\}_{i=1}^{\infty}$.

Definition A.1.1

i) $P_n \equiv \{\text{polynomials of order} \leq n \text{ in the variables } \{\xi_i\}_{i=1}^{\infty}\}$

ii) $P \equiv \bigcup_n P_n$

iii) $\overline{G_n} \equiv \overline{P_n \ominus P_{n-1}} = \overline{\{p \in P_n \mid p \perp P_{n-1}\}}$

$\overline{G_0} = \overline{\{1\}}$

$\overline{G_n}$ is called the n^{th} homogeneous chaos of $\{x(t) \mid t \in T\}$.

Evidently, $\overline{P} = \bigoplus_{n=0}^{\infty} \overline{G_n}$, in which $\oplus$ denotes a direct sum of Hilbert spaces. The $\overline{G_n}$ thus give a Gram-Schmidt type decomposition of $\overline{P}$. To provide orthonormal bases for the $\overline{G_n}$, we introduce the Hermite polynomials, which, as it will be seen are naturally associated to Gaussian processes.

Definition A.1.2 The n^{th} Hermite polynomial is defined as

$$h_n(x) = \frac{(-1)^n}{\sqrt{n!}} e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}$$

Theorem A.1.1 (Homogeneous chaos expansion)

For every n

i) $\{\xi^{\lambda_1 \cdots \lambda_r}_{m_1 \cdots m_r} = h_{m_1}(\xi_{\lambda_1}) \cdots h_{m_r}(\xi_{\lambda_r}) \mid \lambda_1, \cdots, \lambda_r$ pairwise unequal, $m_i > 0\}$

is an orthonormal basis of $\overline{G_n}$.

ii) P is dense in $L^2(\Omega,A,P)$. Thus

$$L^2(\Omega,A,P) = \overline{P} = \overset{\infty}{\underset{n=0}{\oplus}} \overline{G_n}$$

and if $\phi \in L^2(\Omega,A,P)$

$$\phi = \sum_{n=0}^{\infty} \sum_{\{m_1,\cdots,m_r,\lambda_1,\cdots\lambda_r\} \in I_n} (\phi, \varphi_{m_1\cdots m_r}^{\lambda_1\cdots\lambda_r})\, \varphi_{m_1\cdots m_r}^{\lambda_1\cdots\lambda_r}$$

$$I_n \equiv \{m_1,\cdots,m_r,\lambda_1,\cdots\lambda_r \mid m_1+\cdots+m_r=n,\ m_i>0,\ \lambda_1,\cdots,\lambda_r \text{ pairwise unequal}\}$$

APPENDIX 2. SOME CONCEPTS FROM OPERATOR THEORY

The definitions and theorems follow Reed and Simon [41,42].

Let A and B be linear operators on dense domains D and D' respectively in a (complex) Hilbert space H. Let $(\cdot,\cdot)$ denote the inner product in H.

Definition A.2.1

(1) A is *closed* if its graph $\{<\phi,A\phi>|\phi \in D\}$ is a closed subset of $H \times H$ w.r.t. the norm $||<\phi,\psi>|| = ||\phi|| + ||\psi||$.

(2) B extends A if $D \subset D'$ and $B\phi = A\phi$, $\phi \in D$.

(3) $\overline{A|D}$ is the smallest closed extension of A, assuming a closed extension exists.

(4) Spectrum of $A = \sigma(A) = \mathbb{C} - \{\lambda;\ \lambda I - A$ bijects D onto $H\}$.

Definition A.2.2

(1) $D(A^*) = \{\phi \in H|\ \exists \xi \in H$ s.t. $(A\psi,\phi) = (\psi,\xi)\ \forall \psi \in D\}$. For $\phi \in D(A^*)$, $A^*\phi = \xi$.

(2) A is symmetric if $(A\psi,\phi) = (\psi,A\phi)$ for all $\psi,\phi \in D$.

(3) A is self-adjoint if $A = A^*$, that is if A is symmetric and $D(A^*) = D$.

(4) A is essentially self-adjoint if $\overline{A|D}$ is self-adjoint.

Theorem A.2.1 (Spectral theorem)

For A self-adjoint, $\exists!$ P. There is a map λ from the

bounded Borel functions on $\sigma(A)$ to bounded linear operators on H such that

$$\phi(f)\,\phi(g) = \phi(fg)$$

$$\phi(1) = I$$

$$[\phi(f)]^* = \phi[\bar{f}]$$

and if $h_n(x) \to x$ pointwise, $|h_n(x)| \leq |x|$ for all x and n.

$$\lim_{n\to\infty} \phi(h_n)\psi = A\psi, \forall \psi \in D.$$

Example If $-A$ is self-adjoint and bounded below (i.e., $\sigma(-A) \subset (\lambda,\infty)$, $\lambda > -\infty$, $e^{(-t)(-A)} = e^{tA} = \phi(e^{tx})$ is well defined by the spectral theorem for $t \geq 0$. One can show it is a bounded semigroup on H.

Definition A.2.3

$\psi \in D$ is an analytic vector for A if $A^n \psi \in D$ for all n and $\sum_{n=0}^{\infty} \frac{t^n}{n!} \|A^n\psi\|$ has a positive radius of convergence.

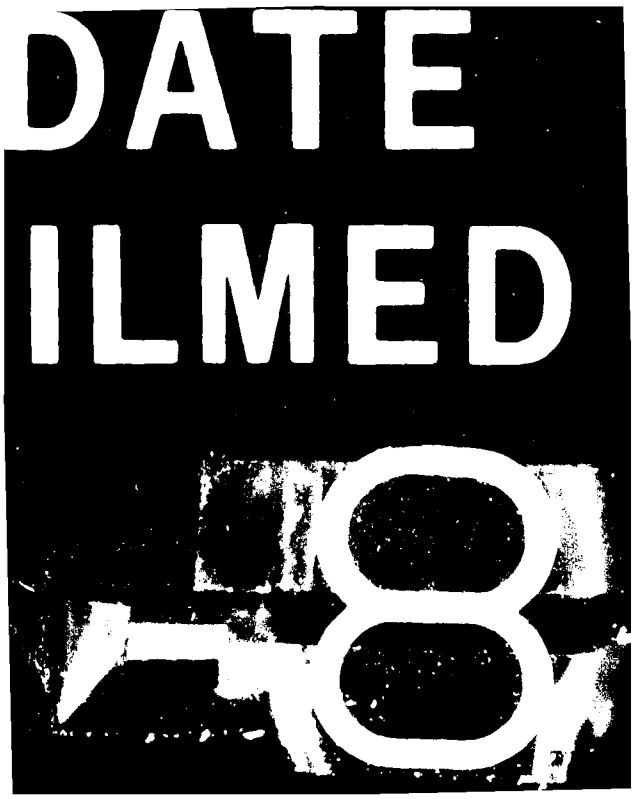
Theorem A.2.2 (Nelson [38]).

Suppose that A *is symmetric and* D *is invariant under* A and contains a dense set of analytic vectors. Then A is essentially self-adjoint on D.

Let G be a finite dimensional Lie algebra, and G its associated simply connected group.

Theorem A.2.3 Let $T(G)$ represent G by skew-symmetric operators on a domain D of H. If there exists $X \in T(G)$ such that iX (on D) is not essentially self-adjoint, then G has no unitary representation on H. Further if $X_1,\cdots,X_d$ is a basis of $T(G)$, $X_1^2+\cdots+X_d^2$ will not be essentially self-adjoint on D either.

For the notion of representation in this theorem see definition 5.1 in chapter 5.


DATE
ILMED